Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations

SpiroMeta Consortium

2019-03

SpiroMeta Consortium , Int COPD Genetics Consortium & Understanding Soc Sci Grp 2019 , ' Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations ' , Nature Genetics , vol. 51 , no. 3 , pp. 494-+ . https://doi.org/10.1038/s41588-

http://hdl.handle.net/10138/313327
https://doi.org/10.1038/s41588-018-0342-2

unspecified
publishedVersion

ARTICLES
https://doi.org/10.1038/s41588-018-0342-2
nature genetics

Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations

Phuwanat Sakornsakolpat[1,2,51], Dmitry Prokopenko[1,3,51], Maxime Lamontagne[4], Nicola F. Reeve[5], Anna L. Guyatt[5], Victoria E. Jackson[5], Nick Shrine[5], Dandi Qiao[1], Traci M. Bartz[6,7,8], Deog Kyeom Kim[9], Mi Kyeong Lee[10], Jeanne C. Latourelle[11], Xingnan Li[12], Jarrett D. Morrow[1], Ma'en Obeidat[13], Annah B. Wyss[10], Per Bakke[14], R. Graham Barr[15], Terri H. Beaty[16], Steven A. Belinsky[17], Guy G. Brusselle[18,19,20], James D. Crapo[21], Kim de Jong[22,23], Dawn L. DeMeo[1,24], Tasha E. Fingerlin[25,26], Sina A. Gharib[27], Amund Gulsvik[14], Ian P. Hall[28,29], John E. Hokanson[30], Woo Jin Kim[31], David A. Lomas[32], Stephanie J. London[10], Deborah A. Meyers[12], George T. O'Connor[33,34], Stephen I. Rennard[35,36], David A. Schwartz[37,38], Pawel Sliwinski[39], David Sparrow[40], David P. Strachan[41], Ruth Tal-Singer[42], Yohannes Tesfaigzi[17], Jørgen Vestbo[43], Judith M. Vonk[22,23], Jae-Joon Yim[44], Xiaobo Zhou[1], Yohan Bossé[4,45], Ani Manichaikul[46,47], Lies Lahousse[18,48], Edwin K. Silverman[1,24], H. Marike Boezen[22,23], Louise V. Wain[5,49], Martin D. Tobin[5,49], Brian D. Hobbs[1,24,52], Michael H. Cho[1,24,52*], SpiroMeta Consortium[50] and International COPD Genetics Consortium[50]

Chronic obstructive pulmonary disease (COPD) is the leading cause of respiratory mortality worldwide. Genetic risk loci provide new insights into disease pathogenesis. We performed a genome-wide association study in 35,735 cases and 222,076 controls from the UK Biobank and additional studies from the International COPD Genetics Consortium. We identified 82 loci associated with $P < 5 \times 10^{-8}$; 47 of these were previously described in association with either COPD or population-based measures of lung function. Of the remaining 35 new loci, 13 were associated with lung function in 79,055 individuals from the SpiroMeta consortium. Using gene expression and regulation data, we identified functional enrichment of COPD risk loci in lung tissue, smooth muscle, and several lung cell types. We found 14 COPD loci shared with either asthma or pulmonary fibrosis. COPD genetic risk loci clustered into groups based on associations with quantitative imaging features and comorbidities. Our analyses provide further support for the genetic susceptibility and heterogeneity of COPD.

COPD is a disease of enormous and growing global burden[1], ranked third as a global cause of death by the World Health Organization in 2016 (ref. [2]). Environmental risk factors, predominately cigarette smoking, account for a large fraction of disease risk, but there is considerable variability in COPD susceptibility among individuals with similar smoking exposure. Studies in families and in populations have demonstrated that genetic factors account for a substantial fraction of disease susceptibility. Similarly to other adult-onset complex diseases, common variants are likely to account for the majority of population genetic susceptibility[3,4]. Our previous efforts have identified 22 genome-wide-significant loci[5]. Expanding the number of loci can lead to new insights into disease pathogenesis, not only through discovery of new biology at individual loci[6,7] but also across loci via identification of functional links and specific cell types and phenotypes[5].

We performed a genome-wide association study (GWAS) combining previously described studies from the International COPD Genetics Consortium (ICGC)[5] with additional subjects from the UK Biobank[8], a population-based study of several hundred thousand subjects with lung function and cigarette smoking assessment. We determined, through bioinformatic and computational analysis, the likely set of variants, genes, cell types, and biological pathways implicated by these associations. Finally, we assessed our genetic findings for relevance to COPD-specific, respiratory, and other phenotypes.

Results

Genome-wide association study of COPD. We included a total of 257,811 individuals from 25 studies in the analysis, including studies from ICGC and UK Biobank (Fig. 1). We defined COPD on the basis of prebronchodilator spirometry according to modified Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria for moderate to very severe airflow limitation[9], as done previously[5]. This definition resulted in 35,735 cases and 222,076 controls (Supplementary Table 1). We tested association of COPD and 6,224,355 variants in a meta-analysis of 25 studies, by using a fixed-effects model. We found no evidence of confounding by population substructure using the linkage disequilibrium score regression[10] (LDSC) intercept (1.0377, standard error (s.e.) = 0.0094).

A full list of affiliations appears at the end of the paper.

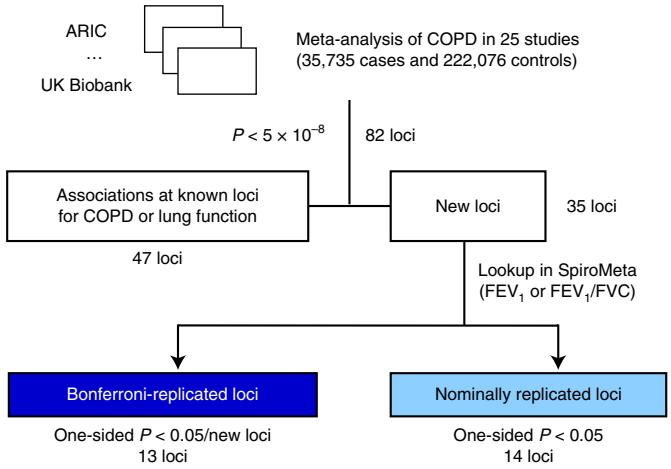


Fig. 1 | Study design. COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in 1 s; FVC, forced vital capacity; ARIC, Atherosclerosis Risk in Communities.

We identified 82 loci (defined by using 2-Mb windows) at genome-wide significance ($P<5\times10^{-8}$) (Figs. 1 and 2, and Supplementary Figs. 1 and 2). Forty-seven of the 82 loci were previously described as genome-wide significant in COPD[5,11] or lung function[12–20] (Supplementary Table 2), thus leaving 35 new loci (Table 1) at the time of analysis. We then sought to replicate these loci. Given the strong genetic correlation between population-based measures of lung function and COPD, we tested the lead variant at each locus for association with forced expiratory volume in 1 s (FEV_1) or FEV_1/forced vital capacity (FVC) in 79,055 individuals from SpiroMeta[21] (Supplementary Table 3). We identified 13 loci—*C1orf87*, *DENND2D*, *DDX1*, *SLMAP*, *BTC*, *FGF18*, *CITED2*, *ITGB8*, *STN1*, *ARNTL*, *SERP2*, *DTWD1*, and *ADAMTSL3*—that replicated by using a Bonferroni correction for one-sided *P* value ($P<0.05/35$; Table 1). Although they did not meet the strict Bonferroni threshold, 14 additional new loci were nominally significant in SpiroMeta (consistent direction of effect and one-sided $P < 0.05$): *ASAP2*, *EML4*, *VGLL4*, *ADCY5*, *HSPA4*, *CCDC69*, *RREB1*, *ID4*, *IER3*, *RFX6*, *MFHAS1*, *COL15A1*, *TEPP*, and *THRA* (Table 1), and all 82 loci showed a consistent direction of effect in COPD and either FEV_1 or FEV_1/FVC ratio in SpiroMeta (Table 1 and Supplementary Table 2). We note that 9 of our 35 new loci were recently described in a contemporaneous analysis of lung function in UK Biobank[21]. None of the new loci appeared to be explained by cigarette smoking, and variant effect sizes in ever- and never-smokers and including and excluding self-reported asthmatics were similar (Supplementary Note). In addition, we found no significant differences in variant effects by sex (Supplementary Note). Including all 82 genome-wide-significant variants, we explained up to 7.0% of the phenotypic variance on the liability scale, using a 10% prevalence for COPD, acknowledging that these effects are likely to be overestimated in the discovery sample. This represents up to a 48% increase in COPD phenotypic variance explained by genetic loci, as compared with the 4.7% explained by the 22 loci reported in a recent GWAS of COPD[5].

Identification of secondary association signals. We used approximate conditional and joint analysis[22] to find secondary signals at each of the 82 genome-wide-significant loci. We found 82 secondary signals at 50 loci, thus resulting in a total of 164 independent associations at 82 loci (Supplementary Table 4). Of the 50 loci containing secondary associations, 33 were at loci previously described for COPD or lung function, and 6 were at Bonferroni-replicated new loci. Of the 82 secondary associations, 20 reached genome-wide significance ($P<5\times10^{-8}$) (Supplementary Table 4). Of the 61 new (not previously described in COPD or lung function) independent associations, 21 reached a region-wise Bonferroni-corrected significance threshold (one-sided $P<0.05$/new independent association(s) at each locus) in unconditioned associations from SpiroMeta (Methods and Supplementary Table 4).

Tissues and specific cell types. In determining the tissue in which COPD genetic variants function to increase COPD risk, the lung is the obvious tissue to consider. However, COPD is a systemic disease[23,24], and within the lung, the cell types collectively contributing to disease pathogenesis are largely unknown. Furthermore, available databases include nonlung cell types, such as smooth muscle in the gastrointestinal (GI) tract, that are relevant to lung biology. To identify putative causal tissues and cell types, we assessed heritability enrichment by using integrated genome annotations at the single-tissue level[25] and tissue-specific epigenomic marks[26]. Lung tissue showed the highest enrichment (enrichment = 9.25, $P=1.36\times10^{-9}$), as previously described, although significant enrichment was also seen in the heart (enrichment = 6.85, $P=3.83\times10^{-8}$) and the GI tract (enrichment = 5.53, $P=6.45\times10^{-11}$). In an analysis of enriched epigenomic marks, the most significant enrichment was in fetal lung and GI smooth muscle DNase-hypersensitivity sites (DHSs) ($P=6.75\times10^{-8}$) and H3K4me1 ($P=7.31\times10^{-7}$) (Supplementary Table 5). To identify the source of association within lung tissue, we tested for heritability enrichment by using single-cell chromatin accessibility[27] (ATAC-seq) and gene expression (RNA-seq) data from human[28,29] and mouse[30] lung (Supplementary Table 5). Using LD score regression in mouse ATAC-seq data, we found enrichment of chromatin accessibility in several cell types, including endothelial cells (most significant) as well as type 1 and type 2 alveolar cells (with the latter among the tissues with the highest fold enrichment; Supplementary Table 5b). Results from using LD score regression[31] or SNPsea[32] on single-cell RNA-seq data varied, with nominal *P* values for genes expressed in type 2 alveolar cells, basal-like cells, club cells, fibroblasts, and smooth muscle cells (Supplementary Table 5c,d).

Fine-mapping of associated loci. To identify the most likely causal variants at each locus, we performed fine-mapping using Bayesian credible sets[33]. Including 160 potential primary and secondary association signals (excluding four variants in the major histocompatibility complex (MHC) region), 61 independent signals had a 99%-credible set with fewer than 50 variants; 34 signals had credible sets with fewer than 20 variants (Supplementary Fig. 3). Eighteen loci had a single variant with a posterior probability of driving association (PPA) greater than 60%, including the *NPNT* (4q24) locus, where the association could be fine-mapped to a single intronic variant, rs34712979 (NC_000004.11:g.106819053 G>A; Supplementary Note and Supplementary Table 6). Most sets included variants that overlapped genic enhancers in lung-related cell types (for example, fetal lung fibroblasts, fetal lung, and adult lung fibroblasts) and were predicted to alter transcription-factor-binding motifs (Supplementary Table 6). Of the 61 credible sets with fewer than 50 variants, 8 sets contained at least one deleterious variant. These deleterious variants included (i) missense variants affecting *TNS1*, *RIN3*, *ADGRG6*, *ADAM19*, *ATP13A2*, *BTC*, and *CRLF3* and (ii) a splice donor variant affecting a long intergenic noncoding RNA (lincRNA), AP003059.2.

Candidate target genes. In most cases, the closest gene to a lead SNP will not be the gene most likely to be the causal or effector gene for disease-associated variants[34–36]. Thus, to identify the potential effector ('target') genes underlying the genetic associations, we integrated additional molecular information including gene expression, gene regulation (open chromatin and methylation data), chromatin

Table 1 | Meta-analysis results showing 35 loci newly associated with COPD and lung function

rsID	HGVS name	Closest gene	Locus	Risk/alt. allele	RAF	UK Biobank			ICGC cohorts			Overall meta-analysis			SpiroMeta FEV_1		SpiroMeta FEV_1/FVC	
						OR	95% CI	*P*	OR	95% CI	*P*	OR	95% CI	*P*	*β*	*P*	*β*	*P*
Significant in SpiroMeta using Bonferroni correction for new loci (one-sided $P < 0.05/35$)																		
rs72673419	NC_000001.10:g.60913143C>T	*C1orf87*	1p32.1	T/C	0.05	1.13	1.07–1.18	2.0×10^{-6}	1.19	1.08–1.31	2.9×10^{-4}	1.14	1.09–1.19	4.0×10^{-9}	-2×10^{-2}	1.3×10^{-1}	-4.7×10^{-2}	3.0×10^{-4}
rs629619	NC_000001.10:g.111738108C>T	*DENND2D*	1p13.3	T/C	0.20	1.08	1.05–1.11	7.5×10^{-8}	1.10	1.04–1.16	8.2×10^{-4}	1.08	1.06–1.11	2.9×10^{-10}	-5.4×10^{-3}	4.6×10^{-1}	-2.4×10^{-2}	9.3×10^{-4}
rs10929386	NC_000002.11:g.15906179C>T	*DDX1*	2p24.3	C/T	0.49	1.06	1.03–1.08	1.2×10^{-6}	1.07	1.03–1.12	1.8×10^{-3}	1.06	1.04–1.08	9.1×10^{-9}	-2.5×10^{-2}	3.1×10^{-6}	-2.4×10^{-2}	7.1×10^{-6}
rs62259026	NC_000003.11:g.57746515C>T	*SLMAP*	3p14.3	C/T	0.75	1.07	1.04–1.09	1.8×10^{-6}	1.07	1.02–1.12	3.8×10^{-3}	1.07	1.04–1.09	2.4×10^{-8}	-2.7×10^{-2}	3.3×10^{-5}	-1.2×10^{-2}	6.4×10^{-2}
rs4585380	NC_000004.11:g.75673363G>A	*BTC*	4q13.3	G/A	0.74	1.07	1.04–1.10	1.2×10^{-7}	1.06	1.02–1.11	8.1×10^{-3}	1.07	1.05–1.09	3.4×10^{-9}	-2.3×10^{-2}	3.7×10^{-4}	-2.0×10^{-2}	1.3×10^{-3}
rs12519165	NC_000005.9:g.170901586A>T	*FGF18*	5q35.1	A/T	0.38	1.08	1.05–1.10	2.7×10^{-10}	1.02	0.97–1.08	4.0×10^{-1}	1.07	1.05–1.09	1.1×10^{-9}	-1.7×10^{-2}	6.5×10^{-3}	-3.1×10^{-2}	2.4×10^{-7}
rs646695	NC_000006.11:g.140280398T>C	*CITED2*	6q24.1	C/T	0.24	1.07	1.05–1.10	3.0×10^{-8}	1.09	1.04–1.14	3.4×10^{-4}	1.08	1.05–1.10	4.6×10^{-11}	-2.2×10^{-2}	5.6×10^{-4}	2.1×10^{-3}	7.5×10^{-1}
rs2040732	NC_000007.13:g.20418134C>T	*ITGB8*	7p21.1	C/T	0.58	1.07	1.05–1.09	5.2×10^{-9}	1.03	0.99–1.07	1.7×10^{-1}	1.06	1.04–1.08	6.9×10^{-9}	-1.6×10^{-2}	1.8×10^{-3}	-1.3×10^{-2}	1.0×10^{-2}
rs1570221	NC_000010.10:g.105656874G>A	*STN1*	10q24.33	A/G	0.35	1.06	1.03–1.08	3.5×10^{-6}	1.07	1.03–1.12	1.4×10^{-3}	1.06	1.04–1.08	2.2×10^{-8}	-1.9×10^{-2}	9.5×10^{-4}	-1.2×10^{-2}	3.9×10^{-2}
rs4757118	NC_000011.9:g.13171236C>T	*ARNTL*	11p15.2	T/C	0.54	1.07	1.04–1.09	1.1×10^{-8}	1.04	1.00–1.08	7.3×10^{-2}	1.06	1.04–1.08	3.8×10^{-9}	-5.7×10^{-3}	2.8×10^{-1}	-1.6×10^{-2}	2.2×10^{-3}
rs9525927	NC_000013.10:g.44842503G>A	*SERP2*	13q14.11	G/A	0.19	1.07	1.04–1.10	2.9×10^{-6}	1.10	1.05–1.16	1.4×10^{-4}	1.08	1.05–1.10	2.8×10^{-9}	-1.5×10^{-2}	2.2×10^{-2}	-2.1×10^{-2}	2.2×10^{-3}
rs72731149	NC_000015.9:g.49984710G>C	*DTWD1*	15q21.2	G/C	0.91	1.12	1.07–1.17	8.9×10^{-7}	1.12	1.04–1.20	2.6×10^{-3}	1.12	1.08–1.16	8.3×10^{-9}	-5.2×10^{-2}	7.9×10^{-7}	-3.5×10^{-2}	9.0×10^{-4}
rs10152300	NC_000015.9:g.84392907G>A	*ADAMTSL3*	15q25.2	G/A	0.23	1.09	1.06–1.11	2.4×10^{-10}	1.08	1.02–1.13	4.7×10^{-3}	1.08	1.06–1.11	4.2×10^{-12}	-1.2×10^{-2}	6.4×10^{-2}	-2.1×10^{-2}	1.5×10^{-3}
Nominally significant in SpiroMeta (one-sided $P < 0.05$)																		
rs955277	NC_000002.11:g.9290357C>T	*ASAP2*	2p25.1	T/C	0.61	1.08	1.05–1.10	2.7×10^{-10}	1.04	0.99–1.08	8.6×10^{-2}	1.07	1.05–1.09	1.9×10^{-10}	-2.9×10^{-4}	9.6×10^{-1}	-1.1×10^{-2}	3.3×10^{-2}
rs12466981	NC_000002.11:g.42433247C>T	*EML4*	2p21	C/T	0.73	1.06	1.03–1.08	8.1×10^{-6}	1.08	1.03–1.13	1.2×10^{-3}	1.06	1.04–1.09	4.9×10^{-8}	-1.2×10^{-2}	4.4×10^{-2}	-1.1×10^{-2}	6.4×10^{-2}
rs2442776	NC_000003.11:g.11640601G>A	*VGLL4*	3p25.3	G/A	0.15	1.09	1.06–1.13	5.7×10^{-8}	1.10	1.04–1.16	8.9×10^{-4}	1.09	1.06–1.12	2.0×10^{-10}	-1.4×10^{-2}	6.4×10^{-2}	-1.1×10^{-2}	1.3×10^{-1}
rs4093840	NC_000003.11:g.123077042T>A	*ADCY5*	3q21.1	A/T	0.47	1.07	1.05–1.09	1.6×10^{-9}	1.04	1.00–1.09	4.8×10^{-2}	1.06	1.04–1.08	3.9×10^{-10}	-9.5×10^{-3}	8.5×10^{-2}	-4.9×10^{-3}	3.8×10^{-1}
rs62375246	NC_000005.9:g.132439010T>A	*HSPA4*	5q31.1	A/T	0.26	1.08	1.05–1.10	9.6×10^{-9}	1.03	0.98–1.08	2.7×10^{-1}	1.06	1.04–1.09	2.2×10^{-8}	-1.5×10^{-2}	1.3×10^{-2}	-7.5×10^{-3}	2.2×10^{-1}
rs979453	NC_000005.9:g.150595073A>G	*CCDC69*	5q33.1	G/A	0.34	1.07	1.05–1.10	1.6×10^{-9}	1.02	0.97–1.06	5.1×10^{-1}	1.06	1.04–1.08	1.4×10^{-8}	-1.1×10^{-2}	7.2×10^{-2}	-2.7×10^{-3}	6.4×10^{-1}
rs1334576	NC_000006.11:g.7211818G>A	*RREB1*	6p24.3	A/G	0.42	1.07	1.05–1.09	4.2×10^{-9}	1.02	0.99–1.06	2.2×10^{-1}	1.06	1.04–1.08	1.2×10^{-8}	2.1×10^{-3}	6.9×10^{-1}	-1.2×10^{-2}	2.9×10^{-2}
rs9350191	NC_000006.11:g.19842661C>T	*ID4*	6p22.3	T/C	0.85	1.11	1.08–1.15	6.2×10^{-12}	1.14	1.05–1.23	1.9×10^{-3}	1.12	1.09–1.15	5.1×10^{-14}	-1.6×10^{-2}	4.9×10^{-2}	-1.1×10^{-2}	1.5×10^{-1}
rs2284174	NC_000006.11:g.30713580T>C	*IER3*	6p21.33	C/T	0.22	1.12	1.10–1.15	3.4×10^{-19}	1.12	1.05–1.21	1.5×10^{-3}	1.12	1.10–1.15	2.1×10^{-21}	-1.4×10^{-2}	5.2×10^{-2}	-1.6×10^{-2}	3.0×10^{-2}
rs674621	NC_000006.11:g.117257018T>C	*RFX6*	6q22.1	C/T	0.32	1.06	1.03–1.08	1.3×10^{-6}	1.07	1.03–1.12	1.4×10^{-3}	1.06	1.04–1.08	7.6×10^{-9}	-1.2×10^{-2}	3.6×10^{-2}	-1.1×10^{-2}	6.9×10^{-2}
rs9329170	NC_000008.10:g.8697658C>G	*MFHAS1*	8p23.1	C/G	0.86	1.10	1.06–1.14	2.0×10^{-8}	1.09	1.03–1.15	4.9×10^{-3}	1.10	1.07–1.13	3.6×10^{-10}	-1.5×10^{-2}	4.6×10^{-2}	-1.4×10^{-2}	6.5×10^{-2}
rs10760580	NC_000009.11:g.101661650G>A	*COL15A1*	9q22.33	G/A	0.71	1.08	1.06–1.11	1.6×10^{-10}	1.04	0.99–1.09	1.1×10^{-1}	1.07	1.05–1.10	1.2×10^{-10}	-1.5×10^{-2}	1.5×10^{-2}	-1.5×10^{-2}	1.4×10^{-2}
rs8044657	NC_000016.9:g.58022625G>A	*TEPP*	16q21	G/A	0.90	1.12	1.07–1.16	1.7×10^{-7}	1.09	1.01–1.17	1.8×10^{-2}	1.11	1.07–1.15	1.1×10^{-8}	-1.6×10^{-2}	9.9×10^{-2}	-2.1×10^{-2}	2.8×10^{-2}
rs62065216	NC_000017.10:g.38218773G>A	*THRA*	17q21.1	A/G	0.42	1.06	1.04–1.09	1.2×10^{-7}	1.06	1.01–1.10	2.2×10^{-2}	1.06	1.04–1.08	8.1×10^{-9}	-9.5×10^{-3}	7.8×10^{-2}	-1.4×10^{-3}	8.0×10^{-1}
Not significant (directionally consistent only)																		
rs4660861	NC_000001.10:g.45946636G>T	*TESK2*	1p34.1	G/T	0.57	1.05	1.03–1.08	7.2×10^{-6}	1.07	1.03–1.12	1.2×10^{-3}	1.06	1.04–1.08	4.4×10^{-8}	-5.5×10^{-3}	3.0×10^{-1}	-4.4×10^{-3}	4.1×10^{-1}
rs7650602	NC_000003.11:g.141147414T>C	*ZBTB38*	3q23	C/T	0.45	1.07	1.04–1.09	6.7×10^{-9}	1.01	0.97–1.06	5.1×10^{-1}	1.06	1.04–1.08	4.9×10^{-8}	1.1×10^{-3}	8.4×10^{-3}	-5.5×10^{-3}	3.2×10^{-1}
rs34651	NC_000005.9:g.72144005C>T	*TNPO1*	5q13.2	C/T	0.08	1.10	1.06–1.15	7.5×10^{-7}	1.12	1.02–1.22	1.2×10^{-2}	1.11	1.07–1.15	3.0×10^{-8}	-3.2×10^{-3}	7.6×10^{-1}	3.4×10^{-2}	7.5×10^{-1}
rs798565	NC_000007.13:g.2752152G>A	*AMZ1*	7p22.3	G/A	0.71	1.09	1.06–1.11	2.8×10^{-11}	1.00	0.95–1.05	8.8×10^{-1}	1.07	1.04–1.09	3.9×10^{-9}	-9.0×10^{-3}	1.2×10^{-1}	-2.9×10^{-3}	6.1×10^{-1}
rs7866939	NC_000009.11:g.85126163T>C	*RASEF*	9q21.32	C/T	0.33	1.06	1.03–1.08	1.3×10^{-6}	1.07	1.02–1.12	3.8×10^{-3}	1.06	1.04–1.08	1.7×10^{-8}	-6.9×10^{-3}	2.5×10^{-1}	-9.5×10^{-3}	1.1×10^{-1}
rs7958945	NC_000012.11:g.115947901A>G	*MED13L*	12q24.21	G/A	0.36	1.06	1.04–1.09	1.0×10^{-7}	1.07	1.02–1.11	2.8×10^{-3}	1.06	1.04–1.09	1.0×10^{-9}	-8.3×10^{-3}	1.5×10^{-1}	-2.9×10^{-3}	6.2×10^{-1}
rs72626215	NC_000019.9:g.46294136G>A	*DMWD*	19q13.32	G/A	0.73	1.06	1.04–1.09	1.7×10^{-6}	1.11	1.05–1.17	7.3×10^{-5}	1.07	1.05–1.10	1.7×10^{-9}	4.2×10^{-3}	4.8×10^{-1}	-9.0×10^{-3}	1.2×10^{-1}
rs73158393	NC_000022.10:g.33335386C>G	*SYN3*	22q12.3	C/G	0.74	1.06	1.04–1.09	1.1×10^{-6}	1.09	1.03–1.14	1.4×10^{-3}	1.07	1.05–1.09	7.7×10^{-9}	-1.8×10^{-3}	7.7×10^{-1}	-5.9×10^{-3}	3.3×10^{-1}

This table shows association statistics in UK Biobank (21,081 cases and 179,711 controls), ICGC cohorts (14,654 cases and 42,365 controls), the overall meta-analysis (35,735 cases and 222,076 controls), and SpiroMeta studies (FEV_1 and FEV_1/FVC; $n = 79,055$). *P* values are two sided, based on Wald statistics (COPD) and *t* statistics (FEV_1 and FEV_1/FVC) without multiple-comparison adjustment. Alt., alternate; RAF, risk allele frequency; CI, confidence interval; OR, odds ratio; ICGC, International COPD Genetics Consortium; HGVS, Human Genome Variation Society.

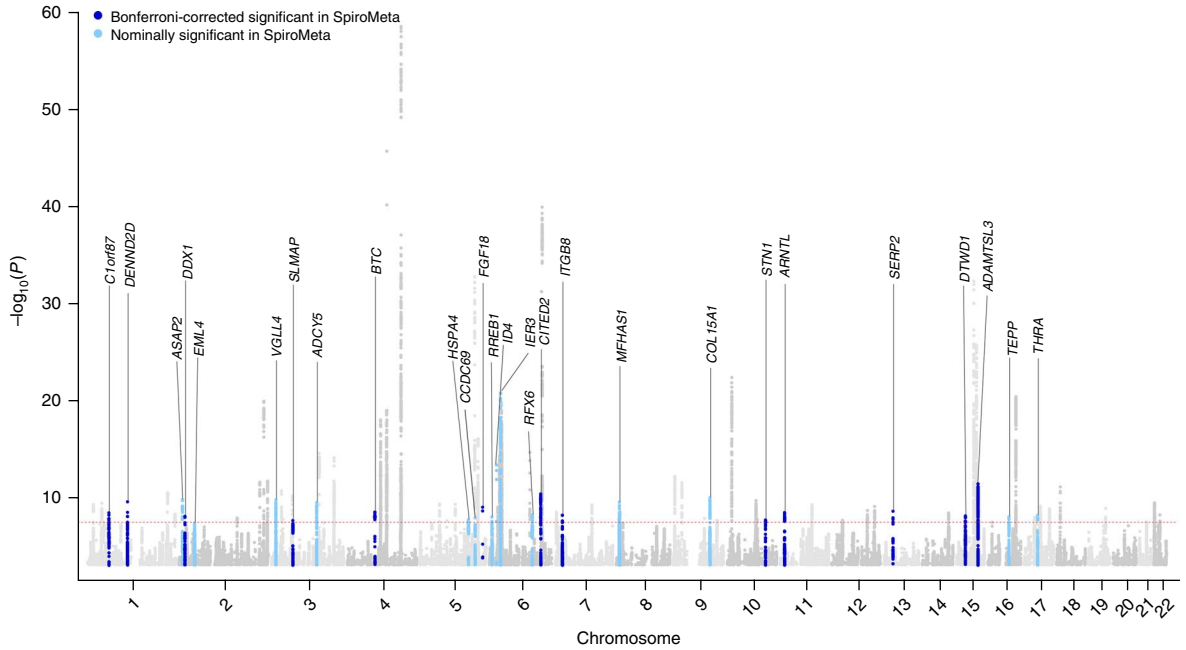


Fig. 2 | Manhattan plot. *P* values are two sided, based on Wald statistics (35,735 cases and 222,076 controls) without multiple-comparison adjustment. Loci are labeled with the closest gene to the lead variant. The red dashed line represents the traditional genome-wide-significance threshold ($P < 5 \times 10^{-8}$).

interaction, co-regulation of gene expression with gene sets, and coding variant data (Fig. 3 and Methods).

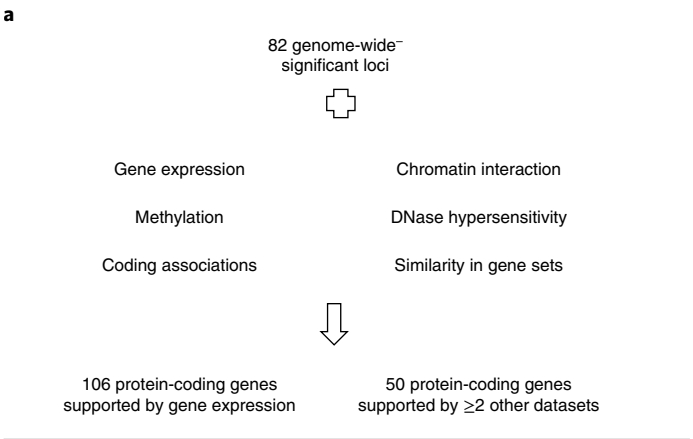
At the 82 associated loci, 472 genes within 1 Mb of the top associated variants were implicated by analysis of at least one dataset; 106 genes were implicated by lung gene expression[37,38], and an additional 50 genes were implicated by two or more other datasets (methylation[39], chromatin interaction[40], open chromatin regions[41], similarity in gene sets[42], or deleterious coding variants[43]; Fig. 3), for a total of 156 genes meeting more stringent criteria. Excluding loci in the MHC region, the median number of potentially implicated genes per locus was four, with a maximum of 17 genes (7q22.1 and 17q21.1). The median distance of the implicated genes to the top associated variants was 346 kb. Of the 82 loci, 60 (73%) included the nearest gene among the implicated genes. We identified 20 genes with supportive evidence from exome-sequencing data. Two genes (*ADAM19* and *ADAMTSL3*) were each implicated by five datasets (Fig. 3), and another two (*EML4* and *RIN3*) were implicated by four datasets. A summary of all genes implicated by using these approaches is included in Supplementary Table 7.

Associated pathways. To gain further functional insight into the associated genetic loci, we performed gene-set enrichment analysis by using DEPICT[42]. Among the 165 enriched gene sets at a false-discovery rate (FDR) <5%, 44% were related to the developmental process term, with a nominal *P* value for lung development of 1.02×10^{-6}; significant subterms included lung alveolus development ($P = 0.0003$) and lung morphogenesis ($P = 0.0005$). We also found enrichment of extracellular-matrix-related pathways, including laminin binding, integrin binding, mesenchyme development, cell–matrix adhesion, and actin filament bundles. Additional pathways of note included histone deacetylase binding, the Wnt-receptor signaling pathway, SMAD binding, the MAPK cascade, and the transmembrane receptor protein serine/threonine kinase signaling pathway. Full enrichment analysis results, including the top genes for each DEPICT gene set, are shown in Supplementary Table 8.

Identification of drug targets. GWAS is also useful for identifying drug targets at either the individual gene[18,44,45] or genome-wide[46,47] level. Of the 472 candidate target genes, 59 genes were targeted by at least one approved or in-development drug[48], for a total of 427 drugs with 134 different modes of action (Supplementary Table 9). Druggable targets at new loci for COPD and lung function included *ABHD6*, *CDKL2*, *GSTO2*, *KCNC4, PDHB, SLK*, and *TRPM7*. We also identified drugs for repurposing in COPD by using transcriptome-wide associations and drug-induced gene expression signatures[49] (Supplementary Note).

Phenotypic effects of COPD-associated variants. To characterize the phenotypic effects of the 82 genome-wide-significant loci, we performed a phenome-wide association analysis within the deeply phenotyped COPDGene study (Methods). We looked for common patterns of phenotype associations for the 82 loci by using hierarchical clustering across scaled *z* scores of phenotype–variant associations. We identified two clusters of variants differentially associated with two sets of phenotypes (Supplementary Fig. 4). Because these two variant–phenotype clusters appeared to be driven by computed tomography (CT) imaging features, we repeated variant clustering but limited analysis to quantitative CT imaging features. We again found two clusters of variants, differentiated by association with quantitative emphysema, emphysema distribution, gas trapping, and airway phenotypes (Fig. 4a). Additionally, we evaluated the association of the 82 genome-wide-significant variants in a prior GWAS of emphysema and airway quantitative CT features[50] (Supplementary Table 10).

We also examined all genome-wide-significant loci in the NHGRI-EBI GWAS Catalog[51] (Supplementary Fig. 5 and Supplementary Table 11) and looked for trait-associated variants in LD ($r^2 > 0.2$) with our lead COPD-associated variants. Many variants were associated with anthropometric measures including height and body mass index (BMI), measurements on blood cells (red and white blood cells), and cancers. COPD is well known to have many common comorbidities, such as coronary artery disease (CAD), type 2 diabetes mellitus (T2D), osteoporosis, and lung cancer. Of these diseases and 13 additional traits, we confirmed previously reported overall genetic correlation (by using LD score regression[52]) of COPD with lung function, asthma, and height and found evidence of a modest



Locus	Gene	GREx	mQTL	Cod.	Hi-C	DHS	GSet
1p13	*DRAM2*	●					
1p13	*OVGP1*		●	●	●	●	
6q24	*HECA*				●	●	
15q25	*ADAMTSL3*	●	●		●	●	●
15q25	*BNC1*	●				●	●
15q25	*HOMER2*					●	●

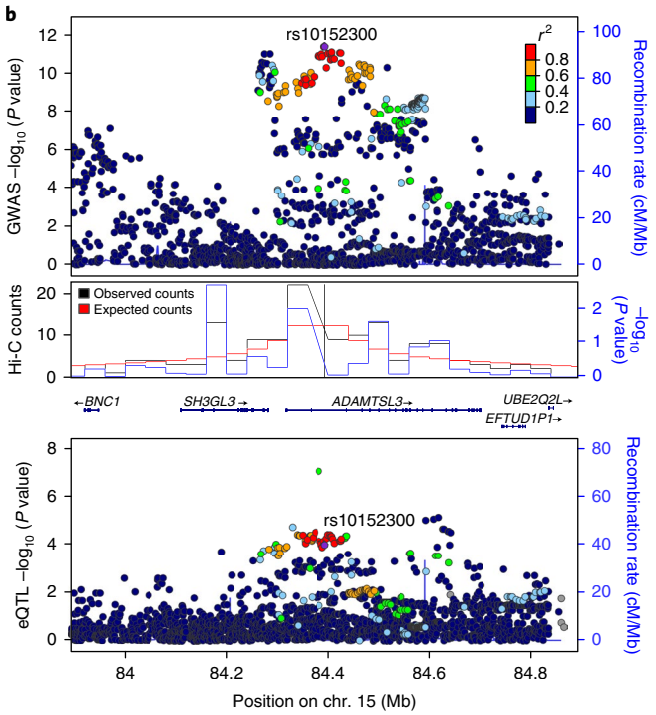


Fig. 3 | Identification of target genes. a, Overview of the datasets used to identify target genes at genome-wide-significant loci. **b**, Regional association plots at the *ADAMTSL3* locus showing GWAS signal (top), chromatin interaction in lung tissue (middle), and expression quantitative trait loci (eQTLs) (bottom). GWAS *P* values are two sided, based on Wald statistics (35,735 cases and 222,076 controls). eQTL *P* values are two sided, based on *t* statistics (1,038 samples). *P* values for Hi-C data were calculated by using binomial distribution from a spline-fitted and outlier-filtered distribution of contacts. No *P* values were adjusted for multiple comparisons. GREx, gene-based association using gene expression; mQTL, colocalization with methylation quantitative trait loci (mQTLs); Cod., significant single-variant or gene-based association tests for deleterious coding variants from exome data; Hi-C, significant chromatin interaction identified in human lung or the IMR90 cell line; DHS, overlap with DHSs; GSet, prioritized genes from DEPICT.

correlation between COPD and lung cancer (Supplementary Note). However, at individual loci and using a more stringent LD threshold ($r^2 > 0.6$), we found evidence of shared risk factors for these comorbid diseases and COPD, including a genome-wide-significant variant near *PABPC4* associated with T2D, four variants associated with CAD (near *CFDP1*, *DMWD*, *STN1*, and *TNS1*), and a variant near *SPPL2C* associated with bone density (Fig. 4b).

Overlapping loci with asthma and pulmonary fibrosis. On the basis of our previous identification of genetic overlap of COPD with asthma, and COPD with pulmonary fibrosis, we examined loci for overlap specifically with these two diseases. In asthma, we noted $r^2 > 0.2$ with one of our variants and previously reported variants at *ID2*, *ZBTB38*, *C5orf56*, *MICA*, *AGER*, *HLA-DQB1*, *ITGB8*, *CLEC16A*, and *THRA*. In pulmonary fibrosis, in addition to our previously described overlap at *FAM13A*, *DSP*, and 17q21, we noted overlapping associations at *ZKSCAN1* and *STN1* (Supplementary Table 12). To more closely examine overlap, we applied a Bayesian method (gwas-pw[53]) to COPD associations from our current GWAS and previous GWAS results for asthma (limited to those of European ancestry) and pulmonary fibrosis[54,55]. To mitigate the effects of including individuals with asthma among our COPD cases, we performed analysis for overlap with asthma after removing self-reported individuals with asthma from UK Biobank for this analysis (Methods). We identified 14 shared genome segments (posterior probability >70%), 9 for asthma and 5 for pulmonary fibrosis (Fig. 4c and Supplementary Table 13). In addition to the three segments shared with pulmonary fibrosis identified in the previous study[5] (*FAM13A*, *DSP*, and the 17q21 locus—here nearest *CRHR1*), we identified two new segments including loci near *ZKSCAN1* and *STN1* (formerly known as *OBFC1*). Shared variants for COPD and pulmonary fibrosis all had opposite directions of effect in the two diseases (that is, increasing risk for COPD but protective for pulmonary fibrosis). In asthma, we identified five shared segments in the 6p21-22 region, as well as the segments containing *ADAM19*, *ARMC2*, *ELAVL2*, and *STAT6*. With the exception of the segment with *STAT6*, overlapping variants showed the same direction of effect.

Discussion

Genetic factors have an important role in COPD susceptibility. We examined genetic risk of COPD in a GWAS of 35,735 cases and 222,076 controls. We identified 82 genome-wide-significant loci for COPD, of which 47 were previously identified in GWAS of COPD or population-based measures of lung function. Of the 35 loci not previously described at the time of analysis, 13 replicated in an independent study of population-based lung function. We used several data sources to attempt to assign causal genes at each locus, identifying 156 genes at the 82 loci that were supported by either gene expression or a combination of at least two other data sources. Our results identify specific genes, cell types, and biological pathways for targeted study and also suggest a genetic basis for the clinical heterogeneity seen in COPD.

Our study supports a role for early life events in the risk of COPD. Gene-set enrichment analysis identified developmental pathways both specific to the lung (for example, lung morphogenesis and lung alveolar development) and related to the lung (for example, the canonical Wnt receptor[56,57] and MAPK–ERK and nerve growth factor receptor signaling pathways). We also confirmed enrichment of heritability for epigenomic marks in the fetal lung. Our findings are consistent with epidemiological studies demonstrating that a substantial portion of the risk for COPD may develop in early

life: genetic variants may set initial lung function[58] and patterns of growth[58–60]. Although further work will be needed to confirm the causal variants and genes affected by our variants, testing the role of these genes in lung-development-relevant mouse or ex vivo models—for example, determining whether perturbation of these genes changes proliferation and differentiation of lung epithelial progenitors in induced pluripotent stem cell (iPSC)-derived alveolar type 2 lung cells[44]—could provide experimental evidence for the role of these genes in early life susceptibility. Ultimately, the goal of this work would be to identify targets for or subsets of high-risk individuals early in the disease course or molecular candidates that may affect lung repair and regeneration[61].

Apart from genes related to lung development, our analyses highlighted several genes and pathways already of interest in COPD therapy (including the COPD target/therapy relationships exemplified by CHRM3/acetylcholine receptor inhibitors and the MAPK pathway/p38 inhibitors)—supporting the role of genetic analyses in finding therapeutic targets[18,62]—and newer genes that may inform future functional studies. We identified *IL17RD* (encoding IL-17 receptor D) as a potential effector gene at the 3p14 locus. Numerous studies have examined the role of IL-17A in COPD[63], and IL17RD can differentially regulate pathways employed by IL-17A[64]. *CHIA* (chitinase acidic) at 1p13.3, which encodes a protein that degrades chitin[65], exhibits lung-specific expression[66,67]. *CHIA* variants have been associated with FEV_1[68], asthma[69–72], and acid mammalian chitinase activity[71,73]. We identified several potential effector genes related to extracellular matrix, cell adhesion, cell–cell interactions, and elastin-associated microfibrils[74–76], some of which have previously been identified in studies of lung function[15]. These include integrin family members that mediate cell–matrix communication (for example, *ITGA1*, *ITGA2*, and *ITGA8*[77–79]), an integrin-ligand-encoding gene (*NPNT*[80]), and genes encoding matrix proteins (for example, *MFAP2* and *ADAMTSL3*). *ADAMTSL3* has a role in cell–matrix interactions related to assembly of fibrillin and microfibril biogenesis[81–83], and of our candidate effector genes, it was supported by the greatest number of bioinformatic analyses. Recombinant forms of other ADAMTS-like proteins have shown evidence in experiments of promoting and enhancing fibrillin and microfibril deposition and assembly[84,85]. *ADAMTSL3* may have a role in preventing emphysematous destruction of lung tissue by ADAMTS in COPD.

In addition to identifying the effector gene for a variant, knowing the effector cell type is critical for functional studies. We identified an overall enrichment of epigenomic marks in lung tissue and GI smooth muscle (also identified in studies of lung function[16]); respiratory smooth muscle was absent from the analyzed datasets. We also performed analyses of single-cell data in an attempt to identify the specific lung cell types in which our top variants are potentially functioning. We found evidence for enrichment of several cell types, including but not limited to endothelial cells, alveolar type 2 cells, and basal-like cells. Each of these cell types has been postulated to have a role in the development of COPD[86–88], and our data are consistent with the likely heterogeneity of lung cell types contributing to COPD susceptibility. The lung comprises at least 40 different resident cell types[89], most of which were not distinctly represented in these datasets. Thus, while our findings support the investigation of specific cell types for further functional studies, they also highlight the need for profiling of lung-relevant cell types and locus-specific analyses.

Characterization of functional variant effects could lead to better disease subtyping and more targeted therapy for COPD. Cluster analysis on hundreds of COPD-associated features in the more extensively phenotyped COPDGene cohort showed heterogeneous effects of genetic variants on COPD-related phenotypes, including CT measurements of airway abnormalities and emphysema—well-described sources of heterogeneity in COPD[90–92]. Analyzing

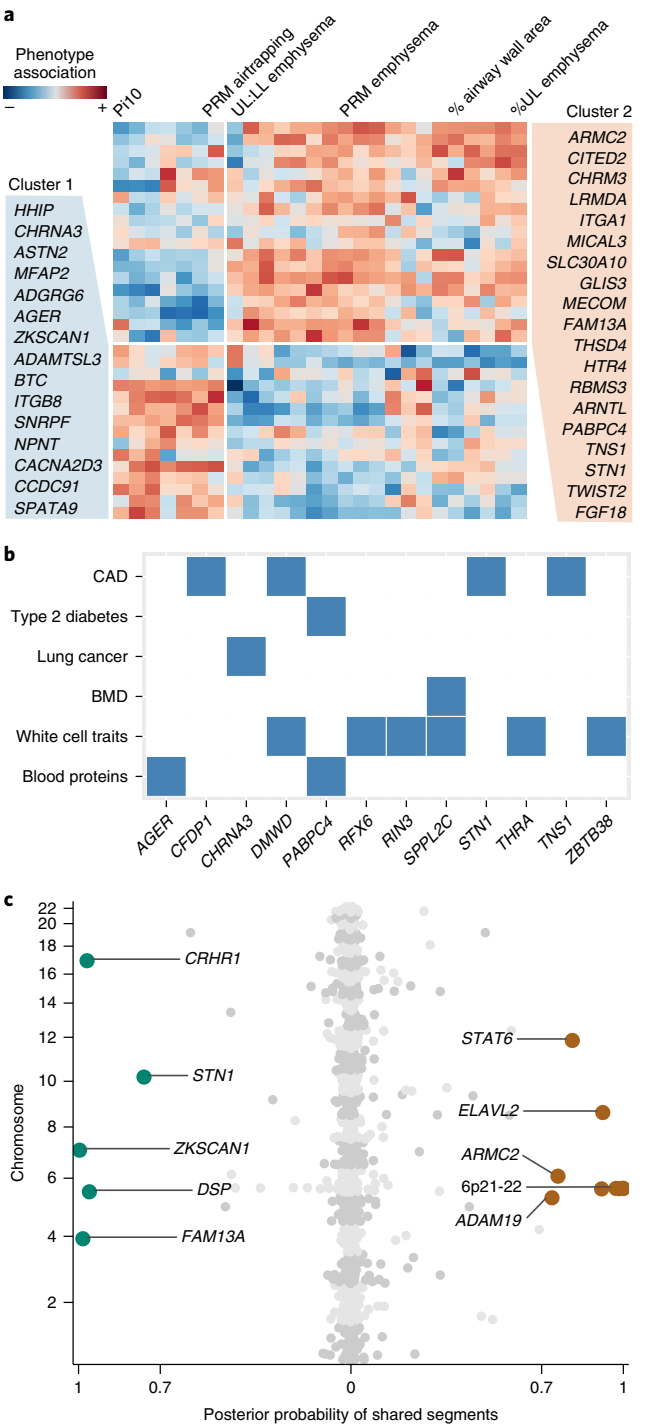


Fig. 4 | Effects on COPD-related and other phenotypes. a, Heat map of scaled CT quantitative imaging associations with the 34 genome-wide-significant variants (known and replicated new associations) with at least nominal ($P < 0.05$) association with any CT imaging feature in COPDGene non-Hispanic white participants. Cluster 1 variants are more associated with airway imaging features, and cluster 2 variants are more associated with emphysema imaging features. Variants are referred to by the closest gene. **b**, Overlap of genome-wide-significant loci for COPD and select traits from the GWAS Catalog. **c**, Genome-wide overlapping results for COPD with pulmonary fibrosis (left) and asthma (right). PRM emphysema, emphysema quantified by parametric response mapping; UL, upper lobe of the lung; LL, lower lobe of the lung; Pi10, airway wall thickness calculated by regressing the square root of the airway wall area with the airway internal perimeter; CAD, coronary artery disease; BMD, bone mineral density.

hundreds of diseases and traits in the GWAS Catalog, we identified overlapping associations with various diseases and traits in multiple organ systems, including comorbidities such as CAD, bone mineral density, and T2D. The COPD-associated *PABPC4* locus was associated with T2D[93] and C-reactive protein (CRP) levels[94]. Although a causal gene in this locus and its contribution to COPD are unknown, the association of this locus with T2D may suggest a shared disease pathway and drug targets. In sum, the identification of variable associations for COPD risk loci with subphenotypes and other diseases[95,96] may have potential for more nuanced approaches to therapy for COPD. Overall, our phenotype, gene, and pathway analyses illustrate the utility of both searching for enrichment of genetic signals overall and performing more detailed identification of the effects of individual variants or groups of variants.

We performed additional specific analysis in two diseases that overlap with COPD, asthma and pulmonary fibrosis. Genome-wide genetic correlation of COPD with asthma has previously been described[5], but our analysis is, to our knowledge, the first to identify specific genetic segments shared by asthma and COPD. While the effects at most of these shared segments were concordant in direction, one of the segments of particular interest near *STAT6* had opposite directions of effect in the two diseases. *STAT6* has a role in helper T cell type 2–dependent inflammation and is activated by IL-4 and IL-13[97]. IL-13 in turn has been found at increased levels in asthmatic airways[98] but at decreased levels in severe emphysema[99]. In pulmonary fibrosis, variants at all overlapping loci have an opposite direction of effect in comparison to COPD[5]. This observation raises the possibility that specific therapies for one disease could increase the risk for the other disease, which may be worth evaluating in treatment trials. The reasons why genetic effects are divergent in COPD and fibrosis are unclear, but these identified opposite effects could point to molecular switches that influence why some smokers develop emphysema whereas others develop pulmonary fibrosis. Although pulmonary fibrosis is an uncommon disease and was specifically excluded in several of our COPD case–control cohorts, interstitial lung abnormalities are increasingly being recognized as a potential precursor to fibrosis, and an inverse relationship between these abnormalities and emphysema has previously been identified[100]. Mechanistically, some have hypothesized that divergent derangement of Wnt and Notch signaling pathways[101] and mesenchymal cell fate[102] may be responsible for the distinct development of these two diseases. We also describe an overlapping region at the *STN1* (previously known as *OBFC1*) locus. *STN1* has a role in telomere maintenance[103]; shortened telomeres have been observed in both COPD and idiopathic pulmonary fibrosis (IPF)[104,105], and rare genetic variants in the telomerase pathway have been implicated in both pulmonary fibrosis and emphysema—though with a concordant direction of genetic risk effects in both diseases[106].

Although our study is a large GWAS of COPD, individuals meeting our criteria for COPD in UK Biobank may be different from those in other studies, especially with respect to smoking history. We used the same definition of COPD as in our previous analysis[5], which included nonsmokers. Our use of prebronchodilator spirometry to define COPD (allowing us to maximize sample size) as well as population-based measures of lung function for replication could bias our findings against variants that are only associated with more severe forms of COPD. We did not exclude other causes of airway obstruction such as asthma, noting that asthma frequently overlaps with and is misdiagnosed in COPD[107]. We performed several additional analyses to determine whether our findings were driven by or markedly different as a result of smoking status, asthma, or use of pre- instead of postbronchodilator spirometry to define COPD. The results of these additional analyses did not indicate a substantial effect of these factors on our overall findings and, together with prior analyses[5,16], suggest that bias due to these factors is likely to be small. However, our study was not designed to identify differences between subgroups, and we cannot rule out a role for studying more severe disease or disease subtypes. We note that the alpha-1 antitrypsin locus (*SERPINA1*) has been identified as genome-wide significant in smaller studies of emphysema and in smokers with severe COPD[108]. In the current study, the PiZ allele (NC_000014.8:g.94844947C>T, rs28929470) was associated with $P = 2.2 \times 10^{-5}$ by using moderate to severe cases (FEV_1 <80% predicted) and with a smaller P value (1.4×10^{-6}) in severe cases (FEV_1 <50% predicted) despite a smaller sample size, a phenomenon that we have previously described[11]. Thus, despite the strong overlap of COPD with quantitative spirometry measures, new loci may be identified through studies of sufficiently large subsets of patients with COPD and with more specific and homogeneous COPD phenotypes. Given the suggestive evidence for replication using a related (but not identical) phenotype for additional new loci beyond the 13 meeting a Bonferroni-corrected threshold for significance, we chose to include all loci significant in discovery in subsequent analyses, recognizing that we probably included some false-positive associations. Our study focused on relatively common variants, predominantly in individuals of European ancestry; more detailed studies of rare variants, the human leukocyte antigen (HLA) regions, and other ancestry groups are warranted, but broader multiancestry analyses are limited by the number of cases in currently available cohorts. Although COPD sex differences have been reported[109], we did not identify significant sex-specific differences in the effect sizes of the 82 top variants. Future studies including more subjects and methodological advances may be needed to elucidate this.

The global burden of COPD is increasing. Our work found a substantial number of new loci for COPD and used multiple lines of supportive evidence to identify potential genes and pathways for both existing and new loci. Further investigation of the genetic overlap of COPD with other respiratory diseases and the phenotypic effects of top loci found new shared loci for asthma and IPF and suggested heterogeneity across COPD-associated loci. Together, these insights provide multiple new avenues for investigation of the underlying biology and potential therapeutics in this deadly disease.

URLs. HGNC database of human gene names, https://www.genenames.org/; Drug Repurposing Hub, https://clue.io/repurposing/; The Query, https://clue.io/query/; Human Genome Region MHC by the Genome Reference Consortium, https://www.ncbi.nlm.nih.gov/grc/human/regions/MHC?asm=GRCh37.p13; EMBL-EBI GWAS Catalog, https://www.ebi.ac.uk/gwas/; LungMAP, https://www.lungmap.net/.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41588-018-0342-2.

Received: 25 June 2018; Accepted: 20 December 2018;
Published online: 25 February 2019

References

1. GBD 2015 Chronic Respiratory Disease Collaborators. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Respir. Med.* **5**, 691–706 (2017).
2. World Health Organization. *Global Health Estimates 2016: Disease burden by Cause, Age, Sex, by Country and by Region, 2000–2016*. https://www.who.int/healthinfo/global_burden_disease/estimates/en/index1.html (2018).
3. Fuchsberger, C. et al. The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47 (2016).

4. Zhou, J. J. et al. Heritability of chronic obstructive pulmonary disease and related phenotypes in smokers. *Am. J. Respir. Crit. Care Med.* **188**, 941–947 (2013).
5. Hobbs, B. D. et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat. Genet.* **49**, 426–432 (2017).
6. Jiang, Z. et al. A chronic obstructive pulmonary disease susceptibility gene, *FAM13A*, regulates protein stability of β-catenin. *Am. J. Respir. Crit. Care Med.* **194**, 185–197 (2016).
7. Lao, T. et al. *Hhip* haploinsufficiency sensitizes mice to age-related emphysema. *Proc. Natl. Acad. Sci. USA* **113**, E4681–E4687 (2016).
8. Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
9. Vogelmeier, C. F. et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. Gold Executive Summary. *Am. J. Respir. Crit. Care Med.* **195**, 557–582 (2017).
10. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
11. Cho, M. H. et al. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir. Med.* **2**, 214–225 (2014).
12. Wilk, J. B. et al. A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet.* **5**, e1000429 (2009).
13. Repapi, E. et al. Genome-wide association study identifies five loci associated with lung function. *Nat. Genet.* **42**, 36–44 (2010).
14. Hancock, D. B. et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat. Genet.* **42**, 45–52 (2010).
15. Soler Artigas, M. et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat. Genet.* **43**, 1082–1090 (2011).
16. Wain, L. V. et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir. Med.* **3**, 769–781 (2015).
17. Soler Artigas, M. et al. Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. *Nat. Commun.* **6**, 8658 (2015).
18. Wain, L. V. et al. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat. Genet.* **49**, 416–425 (2017).
19. Wyss, A. B. et al. Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. *Nat. Commun.* **9**, 2976 (2018).
20. Jackson, V. E. et al. Meta-analysis of exome array data identifies six novel genetic loci for lung function. *Wellcome Open Res.* **3**, 4 (2018).
21. Shrine, N. et al. New genetic signals for lung function highlight pathways and pleiotropy, and chronic obstructive pulmonary disease associations across multiple ancestries. Preprint at https://www.biorxiv.org/content/early/2018/06/12/343293 (2018).
22. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012). S1-3.
23. Agusti, A. & Soriano, J. B. COPD as a systemic disease. *COPD* **5**, 133–138 (2008).
24. Barnes, P. J. & Celli, B. R. Systemic manifestations and comorbidities of COPD. *Eur. Respir. J.* **33**, 1165–1185 (2009).
25. Lu, Q. et al. Systematic tissue-specific functional annotation of the human genome highlights immune-related DNA elements for late-onset Alzheimer's disease. *PLoS Genet.* **13**, e1006933 (2017).
26. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
27. Cusanovich, D. A. et al. A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell* **174**, 1309–1324 (2018).
28. Xu, Y. et al. Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight* **1**, e90558 (2016).
29. Jacob, A. et al. Differentiation of human pluripotent stem cells into functional lung alveolar epithelial cells. *Cell Stem Cell* **21**, 472–488 (2017).
30. Ardini-Poleske, M. E. et al. LungMAP: the molecular atlas of lung development program. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **313**, L733–L740 (2017).
31. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
32. Slowikowski, K., Hu, X. & Raychaudhuri, S. SNPsea: an algorithm to identify cell types, tissues and pathways affected by risk loci. *Bioinformatics* **30**, 2496–2497 (2014).
33. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am. J. Hum. Genet.* **81**, 208–227 (2007).
34. Visscher, P. M. et al. 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).
35. Zhou, X. et al. Identification of a chronic obstructive pulmonary disease genetic determinant that regulates *HHIP*. *Hum. Mol. Genet.* **21**, 1325–1335 (2012).
36. Claussnitzer, M., Hui, C.-C. & Kellis, M. *FTO* obesity variant and adipocyte browning in humans. *N. Engl. J. Med.* **374**, 192–193 (2016).
37. Barbeira, A. N. et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1825 (2018).
38. Lamontagne, M. et al. Leveraging lung tissue transcriptome to uncover candidate causal genes in COPD genetic associations. *Hum. Mol. Genet.* **27**, 1819–1829 (2018).
39. Morrow, J. D. et al. Human lung DNA methylation quantitative trait loci colocalize with COPD genome-wide association loci. *Am. J. Respir. Crit. Care Med.* **197**, 1275–1284 (2018).
40. Schmitt, A. D. et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* **17**, 2042–2059 (2016).
41. Shooshtari, P., Huang, H. & Cotsapas, C. Integrative genetic and epigenetic analysis uncovers regulatory mechanisms of autoimmune disease. *Am. J. Hum. Genet.* **101**, 75–86 (2017).
42. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
43. Qiao, D. et al. Whole exome sequencing analysis in severe chronic obstructive pulmonary disease. *Hum. Mol. Genet.* **27**, 3801–3812 (2018).
44. Sanseau, P. et al. Use of genome-wide association studies for drug repositioning. *Nat. Biotechnol.* **30**, 317–320 (2012).
45. Lencz, T. & Malhotra, A. K. Targeting the schizophrenia genome: a fast track strategy from GWAS to clinic. *Mol. Psychiatry* **20**, 820–826 (2015).
46. Lamb, J. et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 (2006).
47. Sirota, M. et al. Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci. Transl. Med.* **3**, 96ra77 (2011).
48. Corsello, S. M. et al. The Drug Repurposing Hub: a next-generation drug library and information resource. *Nat. Med.* **23**, 405–408 (2017).
49. Subramanian, A. et al. A next generation connectivity map: l1000 platform and the first 1,000,000 profiles. *Cell* **171**, 1437–1452 (2017).
50. Cho, M. H. et al. A genome-wide association study of emphysema and airway quantitative imaging phenotypes. *Am. J. Respir. Crit. Care Med.* **192**, 559–569 (2015).
51. MacArthur, J. et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
52. Zheng, J. et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).
53. Pickrell, J. K. et al. Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**, 709–717 (2016).
54. Demenais, F. et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat. Genet.* **50**, 42–53 (2018).
55. Fingerlin, T. E. et al. Genome-wide imputation study identifies novel HLA locus for pulmonary fibrosis and potential role for auto-immunity in fibrotic idiopathic interstitial pneumonia. *BMC Genet.* **17**, 74 (2016).
56. Skronska-Wasek, W. et al. Reduced frizzled receptor 4 expression prevents WNT/β-catenin-driven alveolar lung repair in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **196**, 172–185 (2017).
57. Sakornsakolpat, P. et al. Integrative genomics identifies new genes associated with severe COPD and emphysema. *Respir. Res.* **19**, 46 (2018).
58. Bui, D. S. et al. Childhood predictors of lung function trajectories and future COPD risk: a prospective cohort study from the first to the sixth decade of life. *Lancet Respir. Med.* **6**, 535–544 (2018).
59. McGeachie, M. J. et al. Patterns of growth and decline in lung function in persistent childhood asthma. *N. Engl. J. Med.* **374**, 1842–1852 (2016).
60. Ross, J. C. et al. Longitudinal modeling of lung function trajectories in smokers with and without COPD. *Am. J. Respir. Crit. Care Med.* **198**, 1033–1042 (2018).
61. Boucherat, O., Morissette, M. C., Provencher, S., Bonnet, S. & Maltais, F. Bridging lung development with chronic obstructive pulmonary disease. Relevance of developmental pathways in chronic obstructive pulmonary disease pathogenesis. *Am. J. Respir. Crit. Care Med.* **193**, 362–375 (2016).

62. Nelson, M. R. et al. The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015).
63. Miossec, P. & Kolls, J. K. Targeting IL-17 and T_H17 cells in chronic inflammation. *Nat. Rev. Drug Discov.* **11**, 763–776 (2012).
64. Mellett, M. et al. Orphan receptor IL-17RD tunes IL-17A signalling and is required for neutrophilia. *Nat. Commun.* **3**, 1119 (2012).
65. O'Leary, N. A. et al. Reference Sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
66. Fagerberg, L. et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* **13**, 397–406 (2014).
67. Saito, A., Ozaki, K., Fujiwara, T., Nakamura, Y. & Tanigami, A. Isolation and mapping of a human lung-specific gene, *TSA1902*, encoding a novel chitinase family member. *Gene* **239**, 325–331 (1999).
68. Aminuddin, F. et al. Genetic association between human chitinases and lung function in COPD. *Hum. Genet.* **131**, 1105–1114 (2012).
69. Birben, E. et al. The effects of an insertion in the 5′UTR of the AMCase on gene expression and pulmonary functions. *Respir. Med.* **105**, 1160–1169 (2011).
70. Chatterjee, R., Batra, J., Das, S., Sharma, S. K. & Ghosh, B. Genetic association of acidic mammalian chitinase with atopic asthma and serum total IgE levels. *J. Allergy Clin. Immunol.* **122**, 202–208 (2008).
71. Ober, C. & Chupp, G. L. The chitinase and chitinase-like proteins: a review of genetic and functional studies in asthma and immune-mediated diseases. *Curr. Opin. Allergy Clin. Immunol.* **9**, 401–408 (2009).
72. Heinzmann, A. et al. Joint influences of acidic-mammalian-chitinase with interleukin-4 and Toll-like receptor-10 with interleukin-13 in the genetics of asthma. *Pediatr. Allergy Immunol.* **21**, e679–e686 (2010).
73. Okawa, K. et al. Loss and gain of human acidic mammalian chitinase activity by nonsynonymous SNPs. *Mol. Biol. Evol.* **33**, 3183–3193 (2016).
74. Yang, J. et al. Rootletin, a novel coiled-coil protein, is a structural component of the ciliary rootlet. *J. Cell Biol.* **159**, 431–440 (2002).
75. Gibson, M. A., Hughes, J. L., Fanning, J. C. & Cleary, E. G. The major antigen of elastin-associated microfibrils is a 31-kDa glycoprotein. *J. Biol. Chem.* **261**, 11429–11436 (1986).
76. Massaro, G. D. et al. Retinoic acid receptor-β: an endogenous inhibitor of the perinatal formation of pulmonary alveoli. *Physiol. Genomics* **4**, 51–57 (2000).
77. Markovics, J. A. et al. Interleukin-1β induces increased transcriptional activation of the transforming growth factor-β-activating integrin subunit β_8 through altering chromatin architecture. *J. Biol. Chem.* **286**, 36864–36874 (2011).
78. Kitamura, H. et al. Mouse and human lung fibroblasts regulate dendritic cell trafficking, airway inflammation, and fibrosis through integrin $\alpha_v\beta_8$-mediated activation of TGF-β. *J. Clin. Invest.* **121**, 2863–2875 (2011).
79. Araya, J. et al. Squamous metaplasia amplifies pathologic epithelial–mesenchymal interactions in COPD patients. *J. Clin. Invest.* **117**, 3551–3562 (2007).
80. Zeltz, C. & Gullberg, D. The integrin–collagen connection: a glue for tissue repair? *J. Cell. Sci.* **129**, 1284 (2016).
81. Hall, N. G., Klenotic, P., Anand-Apte, B. & Apte, S. S. ADAMTSL-3/punctin-2, a novel glycoprotein in extracellular matrix related to the ADAMTS family of metalloproteases. *Matrix Biol.* **22**, 501–510 (2003).
82. Apte, S. S. A disintegrin-like and metalloprotease (reprolysin-type) with thrombospondin type 1 motif (ADAMTS) superfamily: functions and mechanisms. *J. Biol. Chem.* **284**, 31493–31497 (2009).
83. Kutz, W. E. et al. ADAMTS10 protein interacts with fibrillin-1 and promotes its deposition in extracellular matrix of cultured fibroblasts. *J. Biol. Chem.* **286**, 17156–17167 (2011).
84. Gabriel, L. A. R. et al. ADAMTSL4, a secreted glycoprotein widely distributed in the eye, binds fibrillin-1 microfibrils and accelerates microfibril biogenesis. *Invest. Ophthalmol. Vis. Sci.* **53**, 461–469 (2012).
85. Tsutsui, K. et al. ADAMTSL-6 is a novel extracellular matrix protein that binds to fibrillin-1 and promotes fibrillin-1 fibril formation. *J. Biol. Chem.* **285**, 4870–4882 (2010).
86. Ghosh, M. et al. Exhaustion of airway basal progenitor cells in early and established chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **197**, 885–896 (2018).
87. Crystal, R. G. Airway basal cells. The 'smoking gun' of chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **190**, 1355–1362 (2014).
88. Giordano, R. J. et al. Targeted induction of lung endothelial cell apoptosis causes emphysema-like changes in the mouse. *J. Biol. Chem.* **283**, 29447–29460 (2008).
89. Franks, T. J. et al. Resident cellular components of the human lung: current knowledge and goals for research on cell phenotyping and function. *Proc. Am. Thorac. Soc.* **5**, 763–766 (2008).
90. Boschetto, P. et al. Predominant emphysema phenotype in chronic obstructive pulmonary. *Eur. Respir. J.* **21**, 450–454 (2003).
91. Castaldi, P. J. et al. Cluster analysis in the COPDGene study identifies subtypes of smokers with distinct patterns of airway disease and emphysema. *Thorax* **69**, 415–422 (2014).
92. Cerveri, I. et al. The rapid FEV_1 decline in chronic obstructive pulmonary disease is associated with predominant emphysema: a longitudinal study. *COPD* **10**, 55–61 (2013).
93. Bonàs-Guarch, S. et al. Re-analysis of public genetic data reveals a rare X-chromosomal variant associated with type 2 diabetes. *Nat. Commun.* **9**, 321 (2018).
94. Dehghan, A. et al. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation* **123**, 731–738 (2011).
95. Hersh, C. P. et al. Non-emphysematous chronic obstructive pulmonary disease is associated with diabetes mellitus. *BMC Pulm. Med.* **14**, 164 (2014).
96. Higami, Y. et al. Increased epicardial adipose tissue is associated with the airway dominant phenotype of chronic obstructive pulmonary disease. *PLoS One* **11**, e0148794 (2016).
97. Chung, K. F. & Barnes, P. J. Cytokines in asthma. *Thorax* **54**, 825–857 (1999).
98. Kroegel, C., Julius, P., Matthys, H., Virchow, J. C. & Luttmann, W. Endobronchial secretion of interleukin-13 following local allergen challenge in atopic asthma: relationship to interleukin-4 and eosinophil counts. *Eur. Respir. J.* **9**, 899–904 (1996).
99. Boutten, A. et al. Decreased expression of interleukin 13 in human lung emphysema. *Thorax* **59**, 850–854 (2004).
100. Washko, G. R. et al. Lung volumes and emphysema in smokers with interstitial lung abnormalities. *N. Engl. J. Med.* **364**, 897–906 (2011).
101. Chilosi, M., Poletti, V. & Rossi, A. The pathogenesis of COPD and IPF: distinct horns of the same devil? *Respir. Res.* **13**, 3 (2012).
102. Kulkarni, T., O'Reilly, P., Antony, V. B., Gaggar, A. & Thannickal, V. J. Matrix remodeling in pulmonary fibrosis and emphysema. *Am. J. Respir. Cell. Mol. Biol.* **54**, 751–760 (2016).
103. Wan, M., Qin, J., Songyang, Z. & Liu, D. OB fold–containing protein 1 (OBFC1), a human homolog of yeast Stn1, associates with TPP1 and is implicated in telomere length regulation. *J. Biol. Chem.* **284**, 26725–26731 (2009).
104. Albrecht, E. et al. Telomere length in circulating leukocytes is associated with lung function and disease. *Eur. Respir. J.* **43**, 983–992 (2014).
105. Armanios, M. Telomerase and idiopathic pulmonary fibrosis. *Mutat. Res.* **730**, 52–58 (2012).
106. Stanley, S. E. et al. Telomerase mutations in smokers with severe emphysema. *J. Clin. Invest.* **125**, 563–570 (2015).
107. Tinkelman, D. G., Price, D. B., Nordyke, R. J. & Halbert, R. J. Misdiagnosis of COPD and asthma in primary care patients 40 years of age and over. *J. Asthma* **43**, 75–80 (2006).
108. Foreman, M. G. et al. Alpha-1 antitrypsin PiMZ genotype is associated with chronic obstructive pulmonary disease in two racial groups. *Ann. Am. Thorac. Soc.* **14**, 1280–1287 (2017).
109. Han, M. K. et al. Gender and chronic obstructive pulmonary disease: why it matters. *Am. J. Respir. Crit. Care Med.* **176**, 1179–1184 (2007).

Acknowledgements

This work was supported by the Prince Mahidol Award Youth Program Scholarship (P. Sakornsakolpat); NHLBI R01HL084323, R01HL113264, R01HL089856, and P01HL105339 (E.K.S.); K08HL136928 (B.D.H.), the Parker B. Francis Research Opportunity Award (B.D.H.); and R01HL113264, R01HL137927, P01HL105339, and P01HL132825 (M.H.C.). This research was conducted by using the UK Biobank resource under application numbers 20915 (M.H.C.) and 648 (M.D.T.). Please refer to the Supplementary Note for full acknowledgements. Funding bodies had no role in the design of the study, the collection, analysis, or interpretation of the data, or the writing of the manuscript.

Author contributions

P. Sakornsakolpat contributed to study concept and design, data analysis, and manuscript writing. D.P., B.D.H., and M.H.C. contributed to study concept and design, data analysis, statistical support, and manuscript writing. A.B.W., K.d.J., S.J.L., and D.P.S. contributed to study concept and design and to data analysis. P.B., R.G.B., J.D.C., A.G., D.A.M., G.T.O'C., S.I.R., D.A.S., R.T.-S., Y.T., and E.K.S. contributed to study concept and design and to data collection. T.H.B. and J.E.H. contributed to study concept and design and to statistical support. I.P.H., H.M.B., L.V.W., and M.D.T.

contributed to study concept and design. All authors, including those whose initials are not listed above, contributed to critical review and editing of the manuscript and approved the final version of the manuscript.

Competing interests

M.H.C., E.K.S., L.V.W., M.D.T., D.A.L., and I.P.H. have received grant funding from GlaxoSmithKline (GSK). E.K.S. has received honoraria from Novartis for continuing medical education seminars and travel support from GSK. I.P.H. has received grant support from BI. R.T.-S. is an employee and shareholder of GSK. J.V. has received personal fees from GSK, Chiesi Pharmaceuticals, BI, Novartis, and AstraZeneca. D.L.D. has received grants from the National Institutes of Health for research on COPD and personal fees from Novartis. D.A.L. has received honoraria from GSK and chaired the Respiratory Therapy Area Board from 2012 to 2015. Outside the submitted work, L.L. reports expert consultation for Boehringer Ingelheim and Novartis and unrestricted grants from AstraZeneca and Chiesi.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0342-2.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.H.C.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

[1]Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA. [2]Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand. [3]Genetics and Aging Research Unit, Department of Neurology, Massachusetts General Hospital, Boston, MA, USA. [4]Institut Universitaire de Cardiologie et de Pneumologie de Québec, Québec, Quebec, Canada. [5]Genetic Epidemiology Group, Department of Health Sciences, University of Leicester, Leicester, UK. [6]Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA. [7]Department of Medicine, University of Washington, Seattle, WA, USA. [8]Department of Biostatistics, University of Washington, Seattle, WA, USA. [9]Seoul National University College of Medicine, SMG-SNU Boramae Medical Center, Seoul, South Korea. [10]Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Raleigh, NC, USA. [11]Department of Neurology, Boston University School of Medicine, Boston, MA, USA. [12]Department of Medicine, University of Arizona, Tucson, AZ, USA. [13]University of British Columbia Center for Heart Lung Innovation, St. Paul's Hospital, Vancouver, British Columbia, Canada. [14]Department of Clinical Science, University of Bergen, Bergen, Norway. [15]Department of Medicine, College of Physicians and Surgeons and Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA. [16]Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA. [17]Lovelace Respiratory Research Institute, Albuquerque, NM, USA. [18]Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands. [19]Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium. [20]Department of Respiratory Medicine, Erasmus Medical Center, Rotterdam, the Netherlands. [21]Division of Pulmonary and Critical Care Medicine, Department of Medicine, National Jewish Health, Denver, CO, USA. [22]University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, the Netherlands. [23]University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD (GRIAC), Groningen, the Netherlands. [24]Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA. [25]Center for Genes, Environment and Health, National Jewish Health, Denver, CO, USA. [26]Department of Biostatistics and Informatics, University of Colorado Denver, Aurora, CO, USA. [27]Computational Medicine Core, Center for Lung Biology, UW Medicine Sleep Center, Department of Medicine, University of Washington, Seattle, WA, USA. [28]Division of Respiratory Medicine, Queen's Medical Centre, University of Nottingham, Nottingham, UK. [29]National Institute for Health Research Nottingham Biomedical Research Centre, Nottingham, UK. [30]Department of Epidemiology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. [31]Department of Internal Medicine and Environmental Health Center, School of Medicine, Kangwon National University, Chuncheon, South Korea. [32]UCL Respiratory, University College London, London, UK. [33]National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, USA. [34]Pulmonary Center, Department of Medicine, Boston University School of Medicine, Boston, MA, USA. [35]Pulmonary, Critical Care, Sleep and Allergy Division, Department of Internal Medicine, University of Nebraska Medical Center, Omaha, NE, USA. [36]Clinical Discovery Unit, AstraZeneca, Cambridge, UK. [37]Department of Medicine, School of Medicine, University of Colorado Denver, Aurora, CO, USA. [38]Department of Immunology, School of Medicine, University of Colorado Denver, Aurora, CO, USA. [39]2nd Department of Respiratory Medicine, Institute of Tuberculosis and Lung Diseases, Warsaw, Poland. [40]VA Boston Healthcare System and Department of Medicine, Boston University School of Medicine, Boston, MA, USA. [41]Population Health Research Institute, St. George's University of London, London, UK. [42]GlaxoSmithKline R&D, Collegeville, PA, USA. [43]School of Biological Sciences, University of Manchester, Manchester, UK. [44]Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Seoul National University College of Medicine, Seoul, South Korea. [45]Department of Molecular Medicine, Laval University, Québec, Québec, Canada. [46]Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. [47]Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA. [48]Department of Bioanalysis, Ghent University, Ghent, Belgium. [49]National Institute for Health Research Leicester Respiratory Biomedical Research Centre, Glenfield Hospital, Leicester, UK. [50]A list of members and affiliations appears at the end of the paper. [51]These authors contributed equally: Phuwanat Sakornsakolpat, Dmitry Prokopenko. [52]These authors jointly supervised this work: Brian D. Hobbs, Michael H. Cho. *e-mail: remhc@channing.harvard.edu

SpiroMeta Consortium

Nick Shrine[5], Anna L. Guyatt[5], Chiara Batini[5], Jing Hua Zhao[53], Matthias Wielscher[54], Understanding Society Scientific Group[55], Stefan Weiss[56], Katherine A. Kentistou[57,58], James P. Cook[59], Jennie Hui[60,61,62,63], Stefan Karrasch[64,65,66], Medea Imboden[67,68], Sarah E. Harris[69,70], Jonathan Marten[71], Stefan Enroth[72], Shona M. Kerr[71], Ida Surakka[73,74], Veronique Vitart[71], Terho Lehtimäki[75], Ralf Ewert[76], Christian Gieger[77], Georg Homuth[56], Peter K. Joshi[57], Claudia Langenberg[78], Lars Lind[79], Jian'an Luan[78], Anubha Mahajan[80], Alison Murray[81], David J. Porteous[69,70], Rajesh Rawal[77,82], Blair H. Smith[83], Paul R. H. J. Timmers[57], Olli T. Raitakari[84,85], Mika Kähönen[86], Ozren Polasek[57,87], Ulf Gyllensten[72], Igor Rudan[57], Ian J. Deary[69,88], Nicole M. Probst-Hensch[67,68], Holger Schulz[64,66], Alan L. James[60,89,90], James F. Wilson[57,71], Beate Stubbe[76], Eleftheria Zeggini[91,92], Marjo-Riitta Jarvelin[54,93,94,95,96], Nick Wareham[78], Caroline Hayward[71], Andrew P. Morris[59,80], David P. Strachan[41], Ian P. Hall[28,29], Martin D. Tobin[5,49] and Louise V. Wain[5,49]

[53]Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [54]Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment & Health, School of Public Health, Imperial College London, London, UK. [56]Interfaculty Institute for Genetics and Functional Genomics, Department of Functional Genomics, University Medicine Greifswald, Greifswald, Germany. [57]Centre for Global Health Research, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. [58]Centre for Cardiovascular Sciences, Queen's Medical Research Institute, University of Edinburgh, Edinburgh, UK. [59]Department of Biostatistics, University of Liverpool, Liverpool, UK. [60]Busselton Population Medical Research Institute, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia. [61]School of Population Health, University of Western Australia, Crawley, Western Australia, Australia. [62]PathWest Laboratory Medicine of WA, Sir Charles Gairdner Hospital, Crawley, Western Australia, Australia. [63]School of Pathology and Laboratory Medicine, University of Western Australia, Crawley, Western Australia, Australia. [64]Institute of Epidemiology, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany. [65]Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Ludwig-Maximilians-Universitat, Munich, Germany. [66]Comprehensive Pneumology Center Munich (CPC-M), Member of the German Center for Lung Research (DZL), Munich, Germany. [67]Swiss Tropical and Public Health Institute, Basel, Switzerland. [68]University of Basel, Basel, Switzerland. [69]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. [70]Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, UK. [71]Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [72]Department of Immunology, Genetics and Pathology, Uppsala Universitet, Science for Life Laboratory, Uppsala, Sweden. [73]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. [74]National Institute for Health and Welfare (THL), Helsinki, Finland. [75]Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center–Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. [76]Department of Internal Medicine B–Cardiology, Intensive Care, Pulmonary Medicine and Infectious Diseases, University Medicine Greifswald, Greifswald, Germany. [77]Research Unit of Molecular Epidemiology, Institute of Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [78]Medical Research Council Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, UK. [79]Department of Medical Sciences, Cardiovascular Epidemiology, Uppsala University, Uppsala, Sweden. [80]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [81]Institute of Medical Sciences, Aberdeen Biomedical Imaging Centre, University of Aberdeen, Aberdeen, UK. [82]Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany. [83]Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK. [84]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland. [85]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. [86]Department of Clinical Physiology, Tampere University Hospital, and Finnish Cardiovascular Research Center–Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. [87]University of Split School of Medicine, Split, Croatia. [88]Department of Psychology, University of Edinburgh, Edinburgh, UK. [89]Department of Pulmonary Physiology and Sleep Medicine, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia. [90]School of Medicine and Pharmacology, University of Western Australia, Crawley, Western Australia, Australia. [91]Wellcome Sanger Institute, Hinxton, UK. [92]Institute of Translational Genomics, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany. [93]Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland. [94]Biocenter Oulu, University of Oulu, Oulu, Finland. [95]Unit of Primary Health Care, Oulu University Hospital, OYS, Oulu, Finland. [96]Department of Life Sciences, College of Health and Life Sciences, Brunel University London, Uxbridge, UK.

International COPD Genetics Consortium

Alvar Agusti[97], Wayne Anderson[98], Nawar Bakerly[99,100], Per Bakke[14], Robert Bals[101], Kathleen C. Barnes[37], R. Graham Barr[15], Terri H. Beaty[16], Eugene R. Bleecker[12], H. Marike Boezen[22,23], Yohan Bossé[4,45], Russell Bowler[21], Christopher Brightling[49,102], Marleen de Bruijne[103,104], Peter J. Castaldi[1], Bartolome Celli[24], Michael H. Cho[1,24], Harvey O. Coxson[105], James D. Crapo[21], Ron Crystal[106], Pim de Jong[107], Asger Dirksen[108], Jennifer Dy[109], Marilyn Foreman[110], Judith Garcia-Aymerich[111,112,113],

Pierre Gevenois[114], Soumitra Ghosh[42], Hester Gietema[115], Amund Gulsvik[14], Ian P. Hall[28,29], Nadia Hansel[116], Craig P. Hersh[1,24], Brian D. Hobbs[1,24], Eric Hoffman[117], Noor Kalsheker[118], Hans-Ulrich Kauczor[119], Woo Jin Kim[31], Deog Kyeom Kim[9], Tarja Laitinen[120], Diether Lambrechts[121,122], Sang-Do Lee[123], Augusto A. Litonjua[124], David A. Lomas[32], Stephanie J. London[10], Daan W. Loth[18], Sharon M. Lutz[125], David Lynch[126], William MacNee[127], Merry-Lynn McDonald[128], Deborah A. Meyers[12], John D. Newell[117], Borge G. Nordestgaard[129,130], George T. O'Connor[33,34], Ma'en Obeidat[13], Yeon-Mok Oh[123], Peter D. Paré[13,131], Massimo Pistolesi[132], Dirkje S. Postma[23], Milo Puhan[133], Elizabeth Regan[134], Stephen S. Rich[46], Joon Beom Seo[135], Andrea Short[136], Edwin K. Silverman[1,24], David Sparrow[40], Berend Stoel[137], David P. Strachan[41], Nicola Sverzellati[138], Ruth Tal-Singer[42], Gerben ter Riet[139], Yohannes Tesfaigzi[17], Martin D. Tobin[5,49], Edwin J. R. Van Beek[140], Bram van Ginneken[141], Jørgen Vestbo[43], Claus F. Vogelmeier[142], Louise V. Wain[5,49], Adam Wanner[143], George Washko[144], Els Wauters[145], Emiel F. M. Wouters[146], Robert P. Young[147] and Loems Zeigler-Heitbrock[148]

[97]Respiratory Institute, Hospital Clinic, University of Barcelona, IDIBAPS, CIBERES, Barcelona, Spain. [98]Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [99]Salford Royal NHS Foundation Trust, Salford, UK. [100]Manchester Academic Health Sciences Centre, University of Manchester, Manchester, UK. [101]Department of Internal Medicine V–Pulmonology, Allergology and Critical Care Medicine, Saarland University, Homburg, Germany. [102]Department of Respiratory Medicine, Allergy and Thoracic Surgery, Glenfield Hospital, University of Leicester, Leicester, UK. [103]Biomedical Imaging Group Rotterdam, Departments of Radiology and Medical Informatics, Erasmus Medical Center, Rotterdam, the Netherlands. [104]Department of Computer Science, University of Copenhagen, Copenhagen, Denmark. [105]Department of Radiology, University of British Columbia, Vancouver, British Columbia, Canada. [106]Department of Genetic Medicine and Division of Pulmonary and Critical Care Medicine, Department of Medicine, Weill Cornell Medical College, New York, NY, USA. [107]Department of Respiratory Medicine, University Medical Center Utrecht, Utrecht, the Netherlands. [108]Department of Respiratory Medicine, Herlev and Gentofte Hospital, Copenhagen, Denmark. [109]Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA. [110]Morehouse School of Medicine, Atlanta, GA, USA. [111]ISGlobal, Barcelona, Spain. [112]Universitat Pompeu Fabra (UPF), Barcelona, Spain. [113]CIBER Epidemiologia y Salud Publica (CIBERESP), Barcelona, Spain. [114]Department of Radiology, Hôpital Erasme, Université Libre de Bruxelles, Brussels, Belgium. [115]Department of Radiology, Maastricht University Medical Center, Maastricht, the Netherlands. [116]Department of Medicine, Johns Hopkins University, Baltimore, MD, USA. [117]Department of Radiology, University of Iowa, Iowa City, IA, USA. [118]School of Life Sciences, University of Nottingham, Nottingham, UK. [119]Department of Diagnostic and Interventional Radiology, University of Heidelberg, Translational Lung Research Center Heidelberg, Heidelberg, Germany. [120]Department of Pulmonary Diseases and Clinical Allergology, Turku University Hospital, University of Turku, Turku, Finland. [121]Vesalius Research Center (VRC), VIB, Leuven, Belgium. [122]Laboratory for Translational Genetics, Department of Oncology, KU Leuven, Leuven, Belgium. [123]Department of Pulmonary and Critical Care Medicine, and Clinical Research Center for Chronic Obstructive Airway Diseases, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea. [124]Division of Pediatric Pulmonary Medicine, Department of Pediatrics, Golisano Children's Hospital at Strong, University of Rochester Medical Center, Rochester, NY, USA. [125]Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, Boston, MA, USA. [126]Department of Radiology, National Jewish Health, Denver, CO, USA. [127]Department of Respiratory Medicine, University of Edinburgh, Edinburgh, UK. [128]Division of Pulmonary, Allergy and Critical Care Medicine, University of Alabama at Birmingham, Birmingham, AL, USA. [129]Department of Clinical Biochemistry and Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark. [130]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [131]Department of Medicine, University of British Columbia, Institute for Heart and Lung Health, St. Paul's Hospital, Vancouver, British Columbia, Canada. [132]Department of Clinical and Experimental Medicine, University Hospital Careggi, Florence, Italy. [133]Epidemiology, Biostatistics and Prevention Institute (EBPI), University of Zurich, Zurich, Switzerland. [134]Division of Rheumatology, Department of Medicine, National Jewish Health, Denver, CO, USA. [135]Department of Radiology and Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, Songpa-gu, Seoul, Republic of Korea. [136]School of Social Sciences, University of Manchester, Manchester, UK. [137]Leiden University Medical Center, Leiden, the Netherlands. [138]Radiology, Department of Medicine and Surgery, University of Parma, Parma, Italy. [139]Department of General Practice, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands. [140]Edinburgh Imaging, University of Edinburgh, Edinburgh, UK. [141]Department of Radiology and Nuclear Medicine, Radboud University Medical Center, Nijmegen, the Netherlands. [142]Pulmonary Diseases, University of Marburg, Marburg, Germany. [143]Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, Department of Medicine, University of Miami, Miller School of Medicine, Miami, FL, USA. [144]Department of Radiology, Brigham and Women's Hospital, Boston, MA, USA. [145]Laboratory of Pneumology, Department of Chronic Diseases, Metabolism and Ageing, KU Leuven, Leuven, Belgium. [146]Respiratory Medicine, Maastricht University Medical Center, Maastricht, the Netherlands. [147]School of Biological Sciences and Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand. [148]EvA Study Center, Helmholtz Zentrum München, Gauting, Germany.

Methods

Study populations. UK Biobank is a population-based cohort consisting of 502,682 individuals[8]. To determine lung function, we used measures of FEV_1 and FVC derived from spirometry blow-volume time-series data, subjected to additional quality control based on ATS/ERS criteria[110] (Supplementary Note). As in our previous study[5], we defined COPD by using prebronchodilator spirometry according to modified GOLD criteria for moderate to very severe airflow limitation[9]: FEV_1 less than 80% of the predicted value (using reference equations from Hankinson et al.[111]) and a ratio between FEV_1 and FVC of less than 0.7. In agreement with our previous analyses and enrollment criteria for COPD case–control datasets[112], we did not exclude individuals on the basis of self-reported asthma. Genotyping was performed by using the Axiom UK BiLEVE array and the Axiom Biobank array (Affymetrix) and imputed to the Haplotype Reference Consortium (HRC) version 1.1 panel[113].

We invited participants in the prior ICGC COPD GWAS to provide case–control association results (with the exception of the 1958 British Birth Cohort, to avoid overlapping samples with the replication sample). ICGC cohorts performed case–control association analysis based on prebronchodilator measurements of FEV_1 and FEV_1/FVC, and cases were identified by using modified GOLD criteria, as above. Studies were imputed to 1000 Genomes reference panels. Detailed cohort descriptions and cohort-specific methods were previously published[5] (Supplementary Note). All studies complied with all relevant ethical regulations. Ethical/regulatory boards approved the study protocol for each study (Supplementary Note). We obtained informed consent from all participating individuals.

On the basis of the strong genetic overlap of lung function and COPD[5], we performed lookups of select significant variants for FEV_1 and FEV_1/FVC in the SpiroMeta consortium meta-analysis[21]. Briefly, SpiroMeta comprised a total of 79,055 individuals from 22 studies imputed to either the 1000 Genomes Project Phase 1 reference panel (13 studies) or the HRC (9 studies). Each study performed linear regression adjusting for age, age^2, sex, and height, by using rank-based inverse normal transforms, adjusting for population substructure by using principal components or linear mixed models, and performing separate analyses for ever- and never-smokers or using a covariate for smoking (for studies of related subjects). Genomic control was applied to individual studies, and results were combined by using a fixed-effects meta-analysis[21].

Genome-wide association analysis. In UK Biobank, we performed logistic regression of COPD, adjusting for age, sex, genotyping array, smoking pack-years, ever-smoking status, and principal components of genetic ancestry. Association analysis was done by using PLINK 2.0 alpha[114] (downloaded on 11 December 2017) with Firth-fallback settings, using Firth regression when quasicomplete separation or regular-logistic-regression convergence failure occurred. We performed a fixed-effects meta-analysis of all ICGC cohorts and UK Biobank, using METAL (version 2010-08-01)[115]. We assessed population substructure and cryptic relatedness by LD score regression intercept[10]. We defined a genetic locus by using a 2-Mb window (±1 Mb) centered on a lead variant, with conditional analyses as described below.

To maximize our power to identify existing loci and discover new loci, we examined all loci at the genome-wide-significance value of $P < 5 \times 10^{-8}$. We first characterized loci as being previously described (evidence of prior association with lung function[12–20,116,117] or COPD[5,11,118]) or new. We defined previously reported signals if they were in the same LD block in Europeans[119] and in at least moderate LD ($r^2 \geq 0.2$) with previously described loci. For new loci, we attempted replication through association of each lead variant with either FEV_1 or FEV_1/FVC ratio in SpiroMeta, using one-sided P values with Bonferroni correction for the number of new loci examined. New loci failing to meet a Bonferroni-corrected P-value threshold were assessed for nominal significance (one-sided $P < 0.05$) or directional consistency with FEV_1 and FEV_1/FVC ratio in SpiroMeta.

Cigarette smoking is the major environmental risk factor for COPD, and genetic loci associated with cigarette smoking have been reported[5,120]. While we adjusted for cigarette smoking in our analysis, we further examined these effects by additionally testing for association of each locus with cigarette smoking and by looking at two separate analyses of ever- and never-smokers in UK Biobank. We tested for sex-specific genetic effects of genome-wide-significant variants via a stratified analysis and interaction testing, using a 5% Bonferroni-corrected threshold to determine significance (Supplementary Note).

Identification of independent associations at genome-wide-significant loci. We identified specific independent associations at genome-wide-significant loci by using GCTA-COJO[22]. This method uses an approximate conditional and joint analysis approach requiring summary statistics and representative LD information. Because UK Biobank provided the predominant sample, we used 10,000 randomly drawn unrelated individuals from this discovery dataset as an LD reference sample. We scaled genome-wide significance to a 2-Mb region, thus resulting in a locus-wide-significance threshold of 8×10^{-5}, or 2×10^{-6} for variants in the MHC region (chr.6:28,477,797–33,448,354 in hg19). We created regional association plots via LocusZoom, using 1000 Genomes European (EUR) reference data[121] (Nov2014 release).

Identification and prioritization of tissues and cell types, candidate variants, genes, and pathways. *Identification of enriched tissues and specific cell types.* We used LD score regression to estimate enrichment of functional annotations[26] and specifically expressed gene regions[31] on disease heritability. We used LD score regression baseline models (for example, conserved region or promoter-flanking region), tissue-specific annotations from the Roadmap Epigenomics Project[31], integrated tissue annotations from GenoSkyline[25], and cell-type-specific chromatin accessibility data[27] (ATAC-seq). We used four single-cell gene expression (RNA-seq) datasets to identify specific cell types (Supplementary Note), including (i) lung epithelial cells from normal and pulmonary fibrosis human lung[28] (Gene Expression Omnibus (GEO) accession GSE86618), (ii) human iPSC-derived putative alveolar type 2 cells[29] (GSE96642), and (iii) mouse lungs at embryonic day (E) 18.5 and (iv) mouse lungs at postnatal day (P) 1 from Whitsett et al. (unpublished; available at LungMAP[30]). We also used SNPsea[32] to identify enriched cell types in genome-wide-significant loci (Supplementary Note). We report only estimates of coefficients and P values for the Roadmap annotations and gene expression datasets, because these analyses used –h2-cts, which does not report fold enrichment.

Fine-mapping of independent association signals at genome-wide-significant loci. We used Bayesian fine-mapping at each locus to identify the credible set—the set of variants with a 99% probability of containing a causal variant. Briefly, for each genome-wide-significant locus, we calculated the approximate Bayes factors[33] of association. We then selected variants in each locus such that their cumulative posterior probability was greater than or equal to 0.99, by using unscaled variance. At loci with multiple independent associations, we used statistics from approximate conditional analysis with GCTA software on each index variant, adjusting for other independent variants in the locus. Otherwise, we used unconditioned statistics from our meta-analysis. Details on characterization of variant effects are summarized in the Supplementary Note.

Identification of target genes. We used several computational approaches with corresponding available datasets to identify target genes in genome-wide-significant loci. We used two methods that used gene expression data: (i) S-PrediXcan and (ii) DEPICT. We used S-PrediXcan[37] to identify genes with genetically regulated expression associated with COPD. We used data from the Lung-eQTL consortium[38] (1,038 lung tissue samples) as an eQTL and gene expression reference database. S-PrediXcan tests for association between a trait and imputed gene expression by using summary statistics. Here we performed S-PrediXcan, using models for protein-coding genes ±1 Mb from the top associated variants at genome-wide-significant loci. We used DEPICT[42] to prioritize genes from 'reconstituted' gene sets.

We also used additional information on gene regulation, including epigenetic data: (i) regulatory fine-mapping, (ii) mQTLs, and (iii) chromosome-conformation capture. We used regulatory fine-mapping (regfm[41]) to overlap 99%-credible interval variants at each GWAS locus with open chromatin regions, on the basis of DHSs. The DHS cluster accessibility state was then associated with gene expression levels (for 13,771 genes) from 22 tissues in the Roadmap Epigenomics Project[41]. Using both the 99%-credible interval and DHS overlap, as well as DHS state and transcript level association, regfm calculated a posterior probability of association for each gene ±1 Mb from the lead SNP at each GWAS locus. We also searched for overlapping mQTL data from lung tissue, as recently described[39]. To determine whether these signals colocalized (rather than being related owing to LD), we performed colocalization analysis between our GWAS loci and mQTLs in genome-wide-significant loci by using eCAVIAR[122] (Supplementary Note). We also sought information from publicly available chromosome-conformation capture data[40]. We queried association statistics of chromatin contact (that is, long-range chromatin interactions) between top associated variants and gene promoters nearby in the lung (fetal lung fibroblast cell line (IMR90) and human lung tissue[40]) using HUGIn[123]. We retained only the strongest associations (that is, those with the smallest P values) for each cell line or primary cell type in the analysis.

Finally, we searched for signals from deleterious variants by querying the consequences of variants within 99%-credible sets containing fewer than 50 variants (Supplementary Note). We also searched for rare coding variants, on the basis of exome-sequencing results in the COPDGene, Boston Early-Onset COPD (BEOCOPD), and International COPD Genetics Network (ICGN) studies, as previously described[43]. In brief, we performed exome sequencing on 485 severe COPD cases and 504 smoking-resistant controls from the COPDGene study and 1,554 subjects ascertained through 631 probands with severe COPD from the BEOCOPD and ICGN studies. Details on statistical tests for single-variant and gene-based analyses are summarized in the Supplementary Note.

For each dataset described above, we used Bonferroni-corrected P values or a fixed posterior probability threshold to determine target genes at each locus. We report protein-coding genes ±1 Mb from a top associated variant. We restricted our search to genes from the GRCh37 server in biomaRt[124] with updated HUGO Gene Nomenclature Committee (HGNC) names (downloaded from the HGNC database of human gene names on 7 June 2018). For each locus, we used a 5% Bonferroni-corrected threshold (that is, $P < 0.05$ divided by the number of genes at a locus) to determine significance for four data types: gene expression data,

chromatin-conformation capture data, co-regulation of gene expression, and exome-sequencing results. For two remaining datasets, we used a fixed posterior probability (of gene association with a GWAS locus) threshold of 0.1 for regfm and eCAVIAR. We considered genes that were implicated by gene expression or a combination of two or more other datasets (for example, methylation and chromatin-conformation capture data) as target genes.

Identification of pathways. To identify enriched pathways in COPD-associated loci, we performed gene-set enrichment analysis by using the reconstituted gene sets from DEPICT, as described above[42]. We defined significant gene sets using FDR <5%.

Identification of drug targets. We queried our target genes by using the Drug Repurposing Hub[48]. This resource contains comprehensive annotations of launched drugs, drugs in phases 1–3 of clinical development, previously approved and preclinical or tool compounds, curated by using publicly available sources (for example, ChEMBL and DrugBank) and proprietary sources. We performed drug–gene expression similarity analysis[49] (The Query), using a ranked gene set from a gene-based association test[37] (Supplementary Note).

Effects on COPD-related and other phenotypes. COPD is a complex and heterogeneous disorder comprising different biological processes and specific phenotypic effects. In addition, many loci discovered by GWAS have pleiotropic effects. To identify these effects, we performed (i) identification of overlapping genetic loci with related disorders (asthma and pulmonary fibrosis); (ii) genetic association studies of our genome-wide-significant findings by using COPD-related phenotypes, including a cluster analysis to identify groups of variants that might be acting via similar mechanisms; (iii) lookup of top variants in prior COPD-related quantitative CT imaging feature GWAS; (iv) lookup of associations with other diseases and traits by using the GWAS Catalog; and (v) estimation of the genetic correlation between COPD and other diseases and traits.

To identify overlapping loci between COPD and other respiratory disorders, we used gwas-pw[53] to perform pairwise analysis of GWAS. This method searches for shared genomic segments[119] by using an adaptive significance threshold, allowing detection of loci below genome-wide significance. We identified shared segments or variants by using a posterior probability of colocalization greater than 0.7 (ref. [53]). We obtained GWAS summary statistics from previous studies of pulmonary fibrosis[55] and asthma in Europeans[54]. For the overlap analysis of COPD with asthma, we examined the influence of inclusion of individuals with self-reported asthma on both the overlap of discrete GWAS loci (using gwas-pw) and genome-wide genetic correlation (using LD score regression) by performing these analyses in the meta-analysis of ICGC studies and UK Biobank (with individuals with asthma removed from cases in the latter instance). To assess heterogeneous effects of COPD susceptibility loci on COPD-related features (phenotypes), we evaluated associations of our genome-wide-significant SNPs with 121 detailed phenotypes (for example, lung function, CT-imaging-derived metrics, biomarkers, and comorbidities) available for 6,760 COPDGene non-Hispanic white individuals. We calculated z scores for each SNP–phenotype combination relative to the COPD risk allele to create a SNP-by-phenotype z-score matrix. We tested each COPD-related phenotype with at least one nominally significant association with one of our genome-wide-significant COPD SNPs, thus leaving us with 107 phenotypes. We then oriented all z scores to be positive (according to the sign of the median z score) in association with each phenotype to avoid clustering based on the direction of association. To avoid clustering phenotypes only by strength of association with SNPs, we scaled the z scores within each phenotype by subtracting the mean z scores and dividing by the standard deviation of the z scores within each phenotype. We then scaled z scores across SNPs to circumvent clustering of SNPs according to only the relative strength of association with phenotypes. We then performed hierarchical clustering of the scaled z scores of associations between SNPs and phenotypes to identify clusters of SNPs and phenotypes for all 107 phenotypes as well as in the subset of 26 quantitative imaging phenotypes. We performed clustering of variants both in the set of all genome-wide-significant variants in discovery and in the subset of known variants plus new variants meeting a strict Bonferroni threshold in SpiroMeta replication (Supplementary Note). We further examined top variant associations with COPD-related traits through lookup of top variants in a prior GWAS of 12,031 subjects with quantitative emphysema and airway CT features[50]. To examine the overlap of our COPD results with other traits, we downloaded genome-wide-significant associations from the GWAS Catalog[51] ($P < 5 \times 10^{-8}$; downloaded on 10 April 2018). We computed the LD for a pair of COPD- and trait-associated variants within the same LD block in Europeans[119] by using the European-ancestry panel[125] and considered overlap to be present if variants were in at least moderate LD ($r^2 \geq 0.2$). We estimated genetic correlation between COPD and other diseases and traits by using a web engine for LDSC, LD Hub[52]. We assessed the results by using a 5% Bonferroni-corrected significance level.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Genome-wide association summary statistics are available at the database of Genotypes and Phenotypes (dbGaP) under accession phs000179.v5.p2 and via the UK Biobank. Derived phenotypic data for COPD case–control status are also available from UK Biobank.

References

110. Miller, M. R. et al. Standardisation of spirometry. *Eur. Respir. J.* **26**, 319–338 (2005).
111. Hankinson, J. L., Odencrantz, J. R. & Fedan, K. B. Spirometric reference values from a sample of the general U.S. population. *Am. J. Respir. Crit. Care Med.* **159**, 179–187 (1999).
112. Regan, E. A. et al. Genetic epidemiology of COPD (COPDGene) study design. *COPD* **7**, 32–43 (2010).
113. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
114. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
115. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
116. Lutz, S. M. et al. A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. *BMC Genet.* **16**, 138 (2015).
117. Loth, D. W. et al. Genome-wide association analysis identifies six new loci associated with forced vital capacity. *Nat. Genet.* **46**, 669–677 (2014).
118. Hobbs, B. D. et al. Exome array analysis identifies a common variant in *IL27* associated with chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **194**, 48–57 (2016).
119. Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* **32**, 283–285 (2016).
120. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42**, 441–447 (2010).
121. Pruim, R. J. et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
122. Hormozdiari, F. et al. Colocalization of GWAS and eQTL signals detects target genes. *Am. J. Hum. Genet.* **99**, 1245–1260 (2016).
123. Martin, J. S. et al. HUGIn: Hi-C unifying genomic interrogator. *Bioinformatics* **33**, 3793–3795 (2017).
124. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
125. Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555–3557 (2015).

nature research

Corresponding author(s): Michael H. Cho

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	We derived various metrics on spirometry quality control using R. Software for genotype imputation: IMPUTE2 (ARIC, CHS AA, KARE, MESA, SPIROMICS), minimac (CHS EUR, COPDGene, ECLIPSE, FHS, LifeLines, NETT-NAS, GenKOLS), MaCH (COPDGene, ECLIPSE, FHS, NETT-NAS, GenKOLS), and Michigan Imputation Server (eQTL).

Data analysis	We generated genome-wide summary statistics for UK Biobank using PLINK (v.2.0 alpha, https://www.cog-genomics.org/plink/2.0/). We performed a fixed-effects meta-analysis using METAL (v.2010-08-01, http://csg.sph.umich.edu/abecasis/Metal/download/). Regional association plots were generated using LocusZoom (v.1.4, https://github.com/statgen/locuszoom-standalone). We used LDSC (v.1.0, https://github.com/bulik/ldsc/) and SNPsea (v.1.0.3, https://github.com/slowkow/snpsea) to perform enrichment analysis of cell types. Fine-mapping was performed using Wakefield's method (v.2007-04-17, http://faculty.washington.edu/jonno/BFDP.R). We queried variant annotation using BioMart (v.2.36.1, http://www.biomart.org/) and HaploReg (v.4.1, http://archive.broadinstitute.org/mammals/haploreg/haploreg.php). For target gene identification, we tested association of genetically regulated gene expression and COPD using S-PrediXcan (v.0.6.1, https://github.com/hakyimlab/MetaXcan); colocalization analysis using eCAVIAR (v.2.1, https://github.com/fhormoz/caviar); coding associations using SKAT (v.1.3.2.1, https://cran.r-project.org/package=SKAT), GESE (v.2.0.1, https://scholar.harvard.edu/dqiao/gese), GMMAT (v.0.7, https://github.com/lin-lab/GMMAT), and MONSTER (v.1.3, http://www.stat.uchicago.edu/~mcpeek/software/MONSTER/index.html); DNase hypersensitivity sites using regfm (v1.0.0, https://github.com/cotsapaslab/regfm); and reconstituted gene sets using DEPICT (v.140611, https://data.broadinstitute.org/mpg/depict/). We analyzed drug-gene expression signatures using the Query (https://clue.io). Other analyses were done using custom codes in R (v.3.5.0, https://www.r-project.org/).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The genome-wide association summary statistics is available at the database of Genotypes and Phenotypes (dbGaP) under accession phs000179.v5.p2 and via the UK Biobank. Derived phenotypic data for COPD case control status is also available in the UK Biobank.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	UK Biobank is a population-based cohort consisting of 502,682 individuals which was designed to detect genetic effect in multiple traits. The sample size calculation for each trait (e.g., COPD) was not pre-defined. For studies in International COPD Genetics Consortium (ICGC), most studies pre-determined the sufficient sample size according to their designs (e.g., case-control, population-based or family-based studies). Based on the post-hoc sample size calculation, the power to detect genetic association of COPD in UK Biobank was 1 (relative risk=1.1, minor allele frequency=0.3, COPD prevalence=0.1, Ncases=21081, Ncontrols=179713, significance=0.05).
Data exclusions	The exclusion criteria were pre-established. For UK Biobank, we excluded samples based on 1) spirometry quality control (QC) 2) genotyping QC and 3) Genetic ancestry. Detailed descriptions of exclusion criteria were described in Supplementary Methods. Briefly, of 502,682 individuals, 445,754 individuals had at least two measures of FEV1 and FVC and spirometry metrics, age, sex, height, and smoking. We included 324,299 individuals who had reproducible and acceptable measures of spirometry. For genotyping QC, we included 486,367 individuals after excluding outlying heterozygosity or missingness, sex mismatch, >10 3rd degree relatives, and putative sex chromosome aneuploidy. We performed further analyses on individuals with European ancestry identified by K-means clustering. For cohorts part of the International COPD Genetics Consortium, we removed subjects based on phenotype and genotype quality control as previously described. As in UK Biobank, we followed ATS/ERS guidelines for spirometry QC in ICGC studies. As in UK Biobank, ICGC studies followed ATS/ERS guidelines in performing the spirometry QC. We performed genotyping QC using the same rationale as in UK Biobank.
Replication	After we identified novel genome-wide significant associations from the meta-analysis of COPD, we looked up their associations with FEV1 or FEV1/FVC in SpiroMeta, given the high genetic correlation of COPD and lung function. We determined significance in replication using a Bonferroni correction at 5% for a number of novel associations for COPD or lung function, requiring a consistent direction of effect (one-sided Bonferroni-corrected P value < 0.05). We also considered the nominal associations (P < 0.05) and directionally consistent associations. Of 35 novel, 13 loci reached Bonferroni-corrected significance threshold (at P < 0.05) and 14 loci were nominally significant (P < 0.05) in the replication studies in the SpiroMeta. All 35 novel loci were directionally consistent. Loci failing to replicate may be due to the limited sample size in the SpiroMeta consortium.
Randomization	This was observational study. Study individuals were not allocated to experimental groups.
Blinding	Study individuals were not allocated to groups; blinding was not required.

Reporting for specific materials, systems and methods

Materials & experimental systems	
n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods	
n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Characteristics of participants across 25 studies are described fully in Supplementary Table 1. We included adult men and women, smokers and non-smokers. Within the UK Biobank and the majority of the studies in the ICGC, we included individuals of European ancestry. We tested for association between genetic variants and COPD adjusting for covariates including smoking status, and stratification by smoking status.
Recruitment	Participants were recruited based on study design of each individual study (e.g., case-control, population-based, and family-based studies). Informed consent was obtain for all participants and participants for each individual study were recruited in accordance with the standards of the committee or governing body with jurisdiction over human subjects research.